

LINGUISTIC UNIVERSALS AND LANGUAGE CHANGE

Typological rara Synchrony
Natural and unnatural histories
Change explains
Form and function
Structure explains change
Grammaticalization
structure
Analogical change
Paradigm uniformity
Word order
Language contact
Emergent structure
universals
Are there universals?
Economy and adaptation
Selection Diachrony

EDITED BY
JEFF GOOD

LINGUISTIC UNIVERSALS AND LANGUAGE CHANGE

LINGUISTIC UNIVERSALS AND LANGUAGE CHANGE

Edited by

Jeff Good





Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2008 by Oxford University Press

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 978–0–19–929849–5 (Pbk.)
ISBN 978–0–19–922899–7 (Hbk.)

1 3 5 7 9 10 8 6 4 2

CONTENTS

PART II: PHONOLOGICAL UNIVERSALS: VARIATION, CHANGE, AND STRUCTURE

PART III: MORPHOLOGICAL RELATIONSHIPS: THE SHAPE OF PARADIGMS

THE CONTRIBUTORS

Adam Albright received his B.A. in linguistics from Cornell University in 1996 and his Ph.D. in linguistics from UCLA in 2002. He was a Faculty Fellow at the University of California, Santa Cruz from 2002 to 2004, and since then has been an Assistant Professor at the Massachusetts Institute of Technology. His research interests include phonology, morphology, and learnability, with an emphasis on using computational modeling and experimental techniques to investigate issues in phonological theory.

Juliette Blevins is a Senior Scientist in the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology in Leipzig. She received her doctorate in Linguistics from MIT in 1985, and then joined the Department of Linguistics at the University of Texas at Austin. Her research interests range from historical, descriptive, and typological studies, to theoretical analysis with a synthesis in her recent book *Evolutionary Phonology* (2004). Other interests include Oceanic languages, Australian Aboriginal languages, and Native American languages. She is currently working on a sound change database, and a grammar of Yurok, an endangered language of northwestern California.

Joan L. Bybee received her Ph.D. in Linguistics from the UCLA in 1973 and was on the faculty at the University at Buffalo from 1973 to 1989. She is now Distinguished Professor Emerita of Linguistics at the University of New Mexico. Bybee's research interests include theoretical issues in phonology and morphology, language universals, and linguistic change. Her books include *Morphology* (1985), *The Evolution of Grammar* (with Revere Perkins and William Pagliuca, 1994), and *Phonology and Language Use* (2001). Oxford University Press has recently reprinted a collection of her articles under the title *Frequency of Use and the Organization of Language*. In 2004 she served as President of the Linguistic Society of America.

Andrew Garrett received his Ph.D. in 1990 from Harvard University and is Professor of Linguistics at the University of California, Berkeley, where he also serves as Director of the Survey of California and Other Indian Languages. His main theoretical research areas are in historical linguistics, including phonological, morphological, syntactic, and semantic change as well as linguistic diversification and subgrouping. His language research areas include ancient Indo-European languages and Yurok, an indigenous language of northwestern California.

Alice C. Harris received her Ph.D. in linguistics from Harvard University in 1976. She served on the faculty at Vanderbilt University many years, and chaired the Department of Germanic and Slavic Languages there from 1992 until 2002, when she began

teaching at the State University of New York at Stony Brook. Her current research is in synchronic and diachronic problems in morphology; she specializes in languages of the Caucasus, especially Georgian and Udi. Recent books include *Endoclitics and the Origins of Udi Morphosyntax* (2002) and *Historical Syntax in Cross-Linguistic Perspective* (with Lyle Campbell, 1995), winner of the Linguistic Society of America's Bloomfield Book Award in 1998.

Martin Haspelmath is a senior staff member at the Max Planck Institute for Evolutionary Anthropology in Leipzig. He earned degrees at the Universität zu Köln, the University at Buffalo, and the Freie Universität Berlin. He has taught at the Freie Universität Berlin, the Universität Bamberg, the Università di Pavia, the Universität Leipzig, and at summer schools in Albuquerque, Mainz, Düsseldorf, Cagliari, and at MIT. His research interests are primarily in the area of broadly comparative and diachronic morphosyntax (cf. *Indefinite Pronouns*, 1997; *From Space to Time*, 1997; *Understanding Morphology*, 2002). He was one of the editors of Oxford University Press's *World Atlas of Language Structures* (2005).

Bernd Heine is Professor Emeritus of the Institut für Afrikanistik at the Universität zu Köln. He has worked extensively on African languages as well as on grammaticalization. His thirty-three books include *Auxiliaries: Cognitive Forces and Grammaticalization* (1993), *Cognitive Foundations of Grammar* (1997), *Possession: Cognitive Sources, Forces, and Grammaticalization* (1997), and, with Tania Kuteva, *World Lexicon of Grammaticalization* (2002), *Language Contact and Grammatical Change* (2005), *The Changing Languages of Europe* (2006), and *The Genesis of Grammar: A Reconstruction* (2007).

Paul J. Hopper is Paul Mellon Distinguished Professor of the Humanities at Carnegie Mellon University. His publications include *Grammaticalization* (with Elizabeth Closs Traugott, 2003), *A Short Course in Grammar* (1999), *The Limits of Grammaticalization* (co-edited with Anna Giacalone-Ramat, 1998), and *Frequency and the Emergence of Linguistic Structure* (co-edited with Joan L. Bybee, 2001).

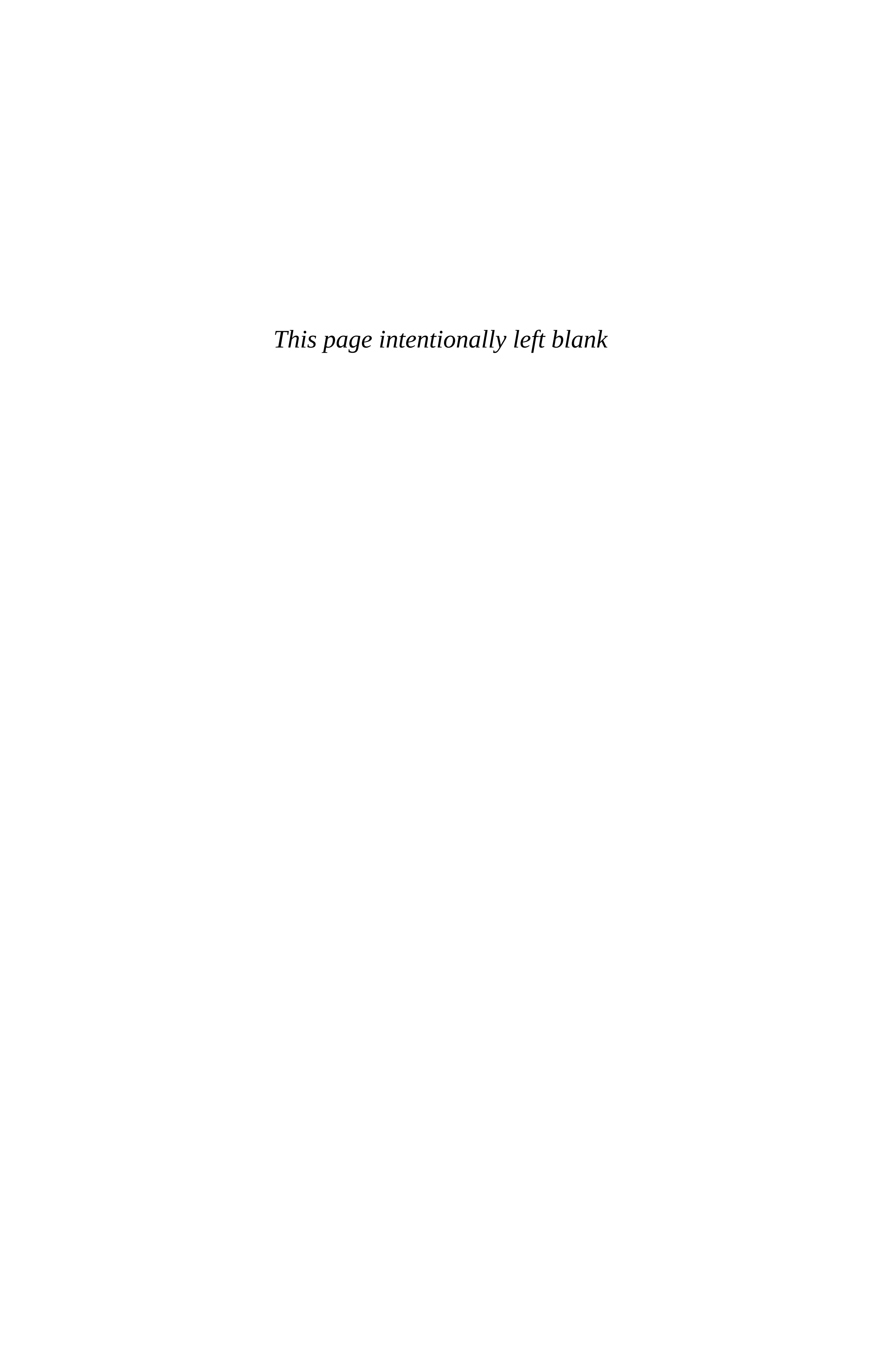
Paul Kiparsky is the Robert M. and Anne T. Bass Professor of Linguistics at Stanford University. He has written on phonology, morphology, historical linguistics, metrics, and the Sanskrit grammatical tradition. His interest in the structure of words and the lexicon is reflected in his writings on Lexical Phonology and Stratal OT, on the relation between morphology, syntax, and thematic roles, and on the principles governing analogical change and grammaticalization.

Tania Kuteva is Professor of English Linguistics at the Institute for English and American Studies, Heinrich-Heine-Universität Düsseldorf. She has taught at a variety of universities worldwide and is author of *Auxiliation: An Enquiry into the Nature of Grammaticalization* (2001) and, with Bernd Heine, *World Lexicon of Grammaticalization* (2002), *Language Contact and Grammatical Change* (2005), *The Changing Languages of Europe* (2006), and *The Genesis of Grammar: A Reconstruction* (2007). In addition, she

is the author of approximately forty articles on grammaticalization, typology, Slavic linguistics, sociolinguistics, and second-language acquisition.

Johanna Nichols is Professor in the Department of Slavic Languages and Literatures, University of California, Berkeley. She works on Slavic and other languages of western Eurasia; Ingush, Chechen, and other languages of the Caucasus; language spreads, especially on the Eurasian steppe; typology; and language prehistory. Her books include dictionaries of Ingush and of Chechen (2004) and *Linguistic Diversity in Space and Time* (1992). She has also written on alignment, case, transitivization/detransitivization, head/dependent marking, and predicate nominals.

John Whitman is Professor and Chair of the Department of Linguistics at Cornell University. He works on syntactic variation and change in a broadly generative framework. He has also written extensively about the synchronic syntax of Japanese and Korean, as well as the history of these languages.



This page intentionally left blank

ABBREVIATIONS

Text

A	Subject of transitive verb
Abs	Absolutive
AGR	Agreement
AP	Adjective phrase
Aux	Auxiliary
C1	First clause
C2	Second clause
CDR	Coargument Disjoint Reference
Cl-Pr	Prosodic Clitic Group
Cobuild	Collins-Birmingham University International Linguistic Database
CP	Complementizer phrase
Comp	Complementizer
CSPAE	Corpus of Spoken American Professional English
DP	Determiner Phrase
E	Event
ECP	Empty Category Principle
FocC	Focused constituent
FocP	Focus phrase
GB	Government and Binding
HMC	Head Movement Constraint
IncE	Incorporated Element
Infl	Inflection
IP	Inflection Phrase
lex	Lexical (i.e., contentful) verb
LF	Logical Form
LFG	Lexical Functional Grammar
LV	Light Verb
Mbh	Mahabharata
MHG	Middle High German
NHG	New High German
N	Noun
NP	Noun phrase
O	Object of transitive verb
OT	Optimality Theory
P	Preposition/Postposition
PP	Prepositional/postpositional phrase
Pr	Prosodic boundary
Pro	Pronoun
pro	null pronoun

PrWd	Prosodic Word
Q	Question
S	subject of intransitive verb
SB-CSAE	Corpus of Spoken American English
SWITCHBOARD	Switchboard Telephone Data Corpus
TETU	The Emergence of the Unmarked
trans	Transitive verb
t	Trace of moved element
UG	Universal Grammar
V	Verb
VP	Verb phrase
X^0-level	Terminal nodes in a syntactic tree
X'	X-bar phrase

Glossing abbreviations

1, 2, 3	First, second, third person
ABL	Ablative
ABS	Absolutive
ACC	Accusative
ADNOM	Adnominal
AGMT	Agreement
ALL	Allative
AOR	Aorist
APPL	Applicative
ASP	Aspect
BA	Chinese element *bà*
CL	Classifier
COMP	Complementizer
CR	Current relevance
DAT	Dative
DEF	Definite
DEM	Demonstrative
EMPH	Emphatic
ERG	Ergative
EV	Evidential
F	Feminine
FACT	Factitive
FUT	Future
GEN	Genitive
HAB	Habitual
HORT	Hortative
IMPF	Imperfect
IMPV	Imperative

IN	Inchoative
IND	Indicative
INF	Infinitive
INSTR	Instrumental
INT	Intensive
INV	Inversion (Georgian)/Inverse (Nocte)
LOC	Locative
M	Masculine
N	Neuter
NAR	Narrative case
NEG	Negative
NOM	Nominative
NONFUT	Non-future tense
OBJ	Object
OBL	Oblique
PASS	Passive
PAST	Past tense
PF	Perfect
PFV	Perfective
PL	Plural
PM	Person marker
POSS	Possessive
PRES	Present
PTCL	Discourse particle
PTCPL	Participle
PV	Preverb
Q	Question marker
REAL	Realis
REFL	Reflexive
REL	Relative marker/pronoun
SBJV	Subjunctive
SBJVI	Udi subjunctive I
SEQ	Sequencer/sequential
SG	Singular
SGT	Singulative
SUBD	Subordinator
SUBJ	Subject
SUP	Supine
TAM	Tense-aspect-mood
TEMP	Temporal
TOP	Topic marker
UNPOSS	Unpossessed
WH	Question word

Work in preparing this volume for publication was done with the assistance of Silke Lambert and Robert Painter and was partially funded by the Julian Park Fund, College of Arts and Sciences, University at Buffalo.

1

Introduction

Jeff Good
University at Buffalo

1.1 DIACHRONY, SYNCHRONY, EXPLANATION, AND UNIVERSALS

Certain grammatical patterns are found again and again in the languages of the world. Some of these patterns recur so frequently that they are given the label "universal". Explaining the source of such patterns is clearly an important goal of linguistics, but how to go about doing this is not obvious. Problems range from the terminological (what sort of patterns should we consider *universal*?), to the methodological (what kind of *explanation* will we accept as sufficient?), to the theoretical (what role does a universal grammar have in shaping recurrent patterns? what role do functional considerations play?). How one answers one of these questions will affect how one answers the others. Can probabilistic generalizations be considered universals? If so, then we need explanations predicting probabilistic patterns. Are we looking for proximate explanations (for example, "language A shows pattern X because it inherited it from its parent language") or ultimate ones (for example, "language A shows pattern X because only this pattern is permitted by Universal Grammar")? Will we assume there is no such thing as Universal Grammar? Then, of course, we cannot appeal to it for any sort of explanation. Will we assume there is such a thing? Then, what is its precise structure?[1]

The papers in this volume are concerned, in one way or another, with both the general problem of explaining recurrent grammatical patterns and the more particular problem of trying to understand what the relationship is between these patterns and language change. Since languages are simultaneously products of history and entities

I would like to thank Adam Albright and several anonymous reviewers for comments on earlier versions of this paper.

[1] Of course, issues like the ones raised by these questions are not applicable solely to the study of language. They could equally well be applied to music, culture, or any other human creation, and many would also apply to the study of biological diversity.

existing at particular times, it seems clear that both diachrony and synchrony have a role to play in explaining the existence of "universals", but where the division of labor lies between the two is contentious. The papers here have been assembled to exemplify a range of approaches to this problem from researchers associated with different subfields of linguistics—e.g., phonology, morphology, syntax—and different approaches to linguistic analysis—e.g., formal, functional, historical. This is not to say the papers themselves (let alone the linguists) fit nicely into these categories. In fact, as we will see, many papers invoke multiple modes of analysis in dealing with the problems they take on.

In this introduction, I will exemplify different approaches to the problem of understanding the relationship between language universals and language change, using the heuristic categories *structural*, *historical*, and *external*.[2] These categories should not be taken as applying to specific researchers or theoretical approaches but rather to modes of analysis as applied to particular problems, and the labels are intended to be partly opaque to avoid any automatic association of these modes of analysis with a given theoretical stance. Section 1.4 summarizes structural approaches, section 1.5 summarizes historical approaches, and section 1.6 summarizes external approaches. Before moving on to those topics, however, first, in section 1.2, I will briefly comment on the term *universal* and, in section 1.3, I will comment on the term *explanation*—as used here.

1.2 ON THE SENSE OF *UNIVERSAL* USED HERE

Just what is a *universal*?[3] Kiparsky's paper makes the interesting distinction between a *typological generalization* and a true universal—the former may represent widely recurrent patterns but fall short of his definition of a universal, which he defines as having to be, among other things, exceptionless. Under such a conception, some classic Greenbergian generalizations would, in fact, be classified as typological generalizations rather than universals—for example, Greenberg's cross-categorial word-order generalizations, discussed in Whitman's contribution.[4]

Bybee's paper, like Kiparsky's, also offers discussion on what sort of patterns ought to be eligible to be considered universals. For her, the true universals are a small set of mechanisms of change. These mechanisms conspire to produce synchronic "universal" patterns—but these patterns themselves are not universal in nature.

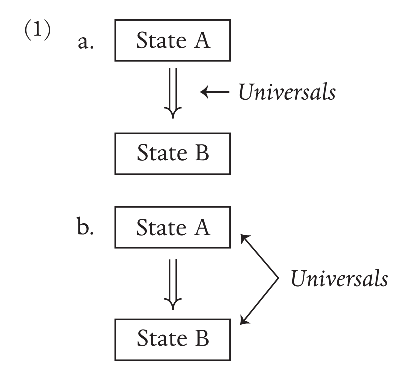
[2] For a recent survey of approaches to language change, see Hickey (2003).

[3] For a recent overview of different senses of and approaches to linguistic universals, see the papers in Mairal and Gil (2006b), in particular Mairal and Gil (2006a).

[4] An example of such a generalization would be a statement like: Languages with SOV basic word order tend to make use of postpositions. That is, they are generalizations about how patterns in one syntactic category correlate with patterns in another category.

While both Bybee and Kiparsky argue that the notion of universal needs to be clearly distinguished from cross-linguistic generalization, the similarities between their approaches largely end there. For Kiparsky, pathways of change that conspire to create common grammatical patterns across languages may be diagnostic of a universal—but they are not themselves universals. True universals, in Kiparsky's view, should be included within a model of the structure of synchronic grammar, a stark contrast to Bybee's position that all true universals are diachronic in nature.

The opposition between Bybee's and Kiparsky's proposals, of course, speaks right to the heart of the theme of a volume whose focus is on the relationship between universals and change. A useful way to schematize the difference between their approaches is given in (1), which makes use of Greenberg's state-process model of language typology (Greenberg 1978a, 1995). This model conceptualizes languages as being in particular synchronic typological states (indicated with boxes), as empirically warranted, and shifting between these states via diachronic processes (indicated with double arrows). Given such a model, Bybee's view of universals can be schematized as in (1a), and Kiparsky's as in (1b).



In general, the other papers in this volume do not take on approaches falling cleanly into either of the schemas in (1). Nevertheless, they provide a convenient way to think about and categorize certain approaches to universals.

While Bybee and Kiparsky are quite explicit about what they consider to be universals, the other authors are less so, and the "universal" patterns they choose to account for are quite varied in nature. Some of the papers focus on grammatical patterns which would probably be considered universals, at least in an informal sense, by most linguists. For example, Whitman examines some Greenbergian word-order generalizations which appear to be exceptionless and would, therefore, be good candidates for the label universal. However, many of the papers veer quite far from the domain of such classical universals. Harris' contribution is the clearest such case. Her focus is not universal patterns but, rather, the

opposite—typologically rare patterns. Of course, any truly complete theory of universal (or even just frequent) patterns in grammar will necessarily also be a theory of rare patterns—making Harris' paper quite relevant here, even if it does not tackle a specific "universal".

The common thread among the chapters is not so much that they attempt to account for any particular universal *per se*. Instead, it is that they are concerned with grammatical patterns which seem to require "universal" explanations—that is, explanations not grounded in the facts of a particular language but which appeal to general principles affecting all languages.

Thus, the motivation behind Kiparsky's examination of how to separate a true universal from a typological generalization, for example, is not simply a call for terminological precision. Rather, he is concerned with identifying what kinds of cross-linguistic patterns should be treated as encoded—in one way or another—in the structure of synchronic grammar (his "true universals") and what kind of patterns can instead be modeled as the result of convergent historical change across languages (his "typological generalizations"). Both types of pattern require explanations applicable to large classes of languages. Nevertheless, the kinds of explanation given to them may need to be quite distinct.

Similarly, Harris is concerned with typologically unusual structures since the fact that they are attested at all means we need to devise a model of grammar which simultaneously predicts they are possible while also explaining why we see them so infrequently—a task which requires a considerably more nuanced approach than simply proposing a model where such structures are excluded entirely.

Surveying the other papers with respect to what kind of "universal" they examine, the two chapters focused on phonology, by Blevins and Bybee, attempt to explain phonological patterns which, if not universal, are certainly prevalent in the world's languages. Specifically, Blevins examines the typology of epenthetic consonants, taking note of common and uncommon kinds of consonant epenthesis, and Bybee looks at the principle of Structure Preservation from a diachronic-functional perspective. Each further argues that the methodological approaches they adopt can be usefully extended to account for a wide range of other phonological patterns.

Albright and Garrett differ from the other contributions by being primarily focused on diachronic universals—specifically, constraints on analogical change. Though they each deal with quite similar data, they adopt strikingly different, though not necessarily contradictory, methodologies. In this volume, these two papers, therefore, most clearly illustrate how a given set of universal patterns can be open to multiple analytical techniques.

Haspelmath's chapter looks at morphosyntactic constructions which, though they may be manifested in very different ways from language to language, tend to show the same asymmetries in coding—for example, a second-person singular imperative form (e.g., *sing!*) typically has a shorter overall form than a

third-person singular "imperative" (e.g., *let her sing!*). While it may be difficult to formulate such patterns in the form of a classical implicational universal, the fact that they recur again and again shows they are clearly in need of a universal explanation.

Kuteva and Heine's contribution is focused on the interaction of language contact with grammaticalization, which allows them to refine our understanding of the many cross-linguistic generalizations uncovered by research on how grammatical morphemes develop. In the introduction to their paper, they also bring up the broader issue that theories of grammaticalization treat the existence of grammatical exceptions, found in all languages, as the "rule", in the sense that such exceptions can often be explained in terms of how the relevant forms came into existence. This sort of "universal"—i.e., the fact that grammars will generally contain subpatterns which run counter to more regular patterns in a language—is a very different sort of cross-linguistic generalization than the type discussed by Kiparsky, but is still clearly in need of a general explanation.

As mentioned above, Whitman's paper focuses on the by-now classic Greenbergian word-order universals. In a similar spirit to Kiparsky's contribution, he revisits them with the goal of determining which patterns can be best understood as resulting from convergent patterns of change and which should be considered as resulting directly from the structure of grammar. Finally, Hopper's paper starts from the premise that, if a certain construction is frequently grammaticalized across unrelated languages, discourse patterns which are the source of the construction should be observable even in languages lacking the construction. In particular, he connects *take* serial verb constructions found in West African languages and Chinese (paraphrasable along the lines of *I take knife cut meat* to mean *I cut the meat with a knife*) to the *take…and* constructions of English (as in, for example, *take this design for the house and enlarge the bedrooms*). He, thus, attempts to show that a pattern which is understood to be part of the "grammar" of many languages may actually exist in other languages in less conspicuous forms—making it, in fact, more universal in nature than might otherwise be supposed.

One theme that emerges, then, from the papers is that "explaining universals" may entail accounting for patterns which would not be considered "universals" in the classic typological sense. That is, if we consider a prototypical synchronic typological universal to take a form like, "If a language has property X, it also has/tends to have property Y" and a prototypical diachronic universal to take a form like, "A language of type A can change directly to a language of type B" (see Greenberg 1995), then many of the papers in this volume do not take "universals" as their central concern at all. Nevertheless, even if they are not concerned with universals in such a narrow sense, the fact that they are all concerned with patterns requiring general explanations clearly makes them concerned with universals in a broader sense.

1.3 *EXPLAINING* UNIVERSALS

What does it mean to *explain* a universal?[5] The explanations offered in this volume are quite varied in nature, both in terms of the strength of the explanation offered and in terms of the causal factors taken as underpinning the explanations. This latter dimension of variation will be the focus of subsequent sections. Here, I will briefly discuss the former.

One kind of explanation we could give for a cross-linguistic pattern would be an *absolute* explanation, which would make (hopefully correct) exceptionless predictions. That is, an explanation which results in a statement like, "This predicts that all languages that have property B will also have property C." Such explanations can be opposed to *probabilistic* explanations which predict when a phenomenon may be likely or unlikely but cannot predict exactly when it will occur. An even weaker type of prediction would simply state the conditions under which a given phenomenon might be found but would have nothing to say about the likelihood of the phenomenon actually appearing under those conditions—an explanation making only this kind of prediction can be labeled *permissive*.

All things being equal, absolute explanations are to be preferred over other kinds, since they make stronger predictions. But, of course, all things are not always equal and our understanding of certain observed patterns at a given time may only permit probabilistic or permissive explanations, and we see all three classes of explanations in the chapters of this volume.

Kiparsky's contribution most clearly exemplifies an absolute explanation for universal patterns (an explanation, however, which is limited only to phenomena meeting his criteria for "true" universals). Specifically, he appeals to a universal grammar constraining the shape of possible human language grammars to explain certain phenomena. A critical factor allowing his explanation to be absolute is that he considers the inclusion of an observable pattern into universal grammar to be contingent upon it meeting a number of criteria, including its being exceptionless. One can of course question the extent to which a universal-grammar–based explanation is "complete" (as Bybee does in her contribution), but this is independent from whether or not the explanation purports to be absolute.

Harris' contribution is quite explicitly a probabilistic one. Her account of typologically unusual structures is not designed to predict exactly when they will or will not occur. Rather, she gives an explanation as to why they should be uncommon in general. Her claim is that such structures are rare because they require a convergence of historical circumstances that is probabilistically unlikely. Importantly, the sort of problem Harris is interested in may, in fact, best be explained

[5] I am grateful to an anonymous reviewer for offering valuable criticisms and insights into the nature of different categories of explanations, many of which are used here—in particular, the categories *absolute*, *probabilistic*, and *permissive*.

probabilistically, and not absolutely. It is clearly possible to give a non-probabilistic account for the existence of a particular unusual structure in a particular language (in fact, Harris does this for two cases in her paper). However, the best general account for the fact that there are grammatical patterns which are attested, but quite rare, may simply be one that has probability at its core—rarity could result from an accidental interaction of independently motivated principles in a given model of grammar and, therefore, be inherently unamenable to an absolute explanation.

Kuteva and Heine's contribution is framed by work done in grammaticalization (see, for example, Heine et al. 1991; Hopper and Traugott 1993) which, often, offers only permissive explanations for phenomena—that is, it focuses on possible grammaticalization paths without, in general, accounting for what factors will cause one language, but not another, to instantiate those paths. In their chapter in this volume, they build on their work integrating contact-induced language change with work on grammaticalization (Heine and Kuteva 2005), allowing them to move towards more probabilistic explanations of certain instances of grammaticalization. This can be seen in, for example, their discussion of "double determination" in Swedish, a label describing a situation where definiteness can be marked by two distinct elements within a noun phrase. They account for the pattern by examining both independently exemplified grammaticalization pathways and the areal patternings of definite marking in Scandinavian languages, thereby allowing them to explain why this grammaticalization pattern is found in some dialects but not others.

Most of the papers in this volume offer probabilistic explanations, as opposed to absolute or permissive ones, because they are attempting to explain generalizations which are themselves probabilistic. This is true, for example, of Blevins' account of common and uncommon epenthesis phenomena, Haspelmath's discussion of coding asymmetries, Hopper's discussion of *take* "serial verb" constructions, and (some of) Whitman's discussion of word-order correlations. Bybee's paper is similar in this regard in its account for Structure Preservation, a principle once proposed as describing an absolute generalization but which is now known to have exceptions.

Two of the other papers in this volume, Albright's and Garrett's, look at directionality of analogical change, which is amenable to being characterized in terms of absolute generalizations—and, therefore, to being given absolute explanations. (Of course, predicting whether or not a given possible analogical change actually will or will not occur would seem to be more problematic in this regard.) Garrett argues that paradigm-leveling is *always* the result of imposition of one paradigmatic pattern on another. This constitutes a "low-level" absolute explanation for possible directions of analogical change. Albright looks at similar data and offers a different absolute explanation for the directions of analogical change, based on the idea that analogical change will extend a base which more reliably predicts an entire paradigm over a base which is less reliable. While Albright's approach makes use of statistical information

in determining reliability, it uses this information to, in fact, make absolute, not probabilistic, predictions about possible directions of analogical change.

In addition to the question of how powerful an explanation is, there is another dimension of explanation worth briefly discussing here: whether a given explanation is in terms of a proximate cause or an ultimate one. The nature of this distinction comes through clearly in the current selection of papers in Bybee's contribution. While the details of her account of Structure Preservation are given in terms of a specific scenario for the development of word-level contrasts (a proximate explanation), this account is situated within a broader framework which seeks to explain how language use, in general, gives rise to language structure. Furthermore, the general usage-based principles Bybee gives could themselves be grounded in more general neuro-cognitive principles. This would, at least from the perspective of the linguist, allow for an ultimate explanation by giving a non-linguistic explanation for proposed linguistic principles.

Haspelmath's paper offers a comparable example, giving both a proximate and an ultimate account for morphosyntactic coding asymmetries. On one level, they are explained as arising through differential patterns of change—that is, their proximate cause is taken to be historical in nature. However, these patterns of change are themselves explained as a result of the fact that human beings generally act purposefully and rationally—an ultimate cause which plays out in language through diachronic mechanisms, along the lines of the Invisible Hand model developed by Keller (1994). Hopper's paper, too, offers both a proximate and an ultimate explanation for the phenomenon he focuses on, insofar as his account of the development of a specific English construction is grounded in general rhetorical principles.

Deciding whether a given explanation in this volume may be a proximate one or an ultimate one is not always straightforward and can hinge upon, among other things, one's theoretical inclinations. A generativist may consider an explanation invoking the "structure" of (universal) grammar to be an ultimate one, while a functionalist may see such an explanation as merely a convenient stopping point en route to a "deeper" explanation.

In addition to these non-linguistic parameters of "explanation", there is also, of course, the issue of what linguistic principles are invoked to explain a given phenomenon. The set of such allowed principles, of course, has been a topic of great interest in the generative era, debated by both generativists and non-generativists. The collection of papers in Hawkins (1988a), for example, offers a diversity of viewpoints in this area, with Hawkins (1988b) providing a useful summary of work done to that point. The next sections of this introduction will include discussion of approaches to the explanation of language universals which are of specific relevance to the theme of the present volume: those focusing on the relationship between universals and change. Three heuristic categories of linguistic explanations will be discussed, *structural*, *historical*, and *external*, each of which is taken up in turn.

1.4 STRUCTURAL APPROACHES

Structural approaches to universals claim that a particular universal can be explained as the result of inherent, universal aspects of grammar. Just what would constitute the requisite "universal grammar" is, of course, a matter of debate, but how that debate may be resolved is an independent matter from the idea that universals can be fruitfully explained by appealing to the shape of an abstract universal linguistic structure. If one accepts this, then, as Kiparsky aptly puts it in his contribution to this volume, one must also accept that "synchronic assumptions have diachronic consequences". This idea was schematically represented above in (1b), where universals were treated as being applicable to states of languages and not to the processes through which languages transition from one state to another.

The idea that synchronic structure may, in some sense, explain the nature of change is hardly new. Kiparsky points this out in his contribution, citing, among other instances, Ferdinand de Saussure's explanation for the regularity of sound change. "Sound change, as we have seen ... affects not words, but sounds (de Saussure 1916/2005: 143)." That is, sound change affects any signifier containing the relevant sound, regardless of what signs the signifier is part of de Saussure's view of sounds and concepts as being two independent facets of the structure of the synchronic sign, therefore, was the basis for his explanation for the regularity of sound change.

Methodologically, we can distinguish between two types of structural approaches to the relationship between diachrony and universals. The first type is well exemplified by King (1969), which is "generally aimed at developing a theory of change which could hook up to the existing synchronic theory, so as to correctly characterize the possible forms of linguistic change, and the constraints to which they are subject (Kiparsky 1982: 57 (originally published as Kiparsky 1978))."[6] The second is more concerned with using diachronic facts to help refine synchronic models of structure. Kiparsky (1982: 57) exemplifies this sort of work quite well, with statements like, "[T]he present state of linguistics is such that the synchronic theory is often rather indeterminate in exactly the respects that would be most relevant for historical linguistics. For this reason much progress in historical linguistics depends on sharpening synchronic theory so that it will provide the right basis for diachronic explanation." Such work, focusing on the interplay between synchronic and diachronic data in developing structural models, is clearly important in the present context.

[6] In syntax, the majority of the papers in Pintzuk et al. (2000) and Lightfoot (2002), two recent volumes, would fall into this line of work. In morphology, as discussed by Garrett in his chapter, constraints have been proposed to account for paradigm uniformity synchronically. Such constraints could clearly also be used to account for historical change. Hock (1991), though not explicitly attempting to unify a synchronic theory of phonology with diachronic change, makes use of a set of distinctive features in describing many sound changes, which would seem to put him into a similar category as King. Extensive discussion of different kinds of explanations for phonological change, including synchronically oriented accounts, can be found in Blevins (2004a), which does the topic far more justice than I can here.

The collection of papers in Kiparsky (1982) brings together a number of arguments for a structural approach to the relationship between universals and change in the realm of phonology. In syntax, the work of Anthony Kroch and his associates is also noteworthy in this regard. In a series of papers including Kroch (1989a, 1989b), Kroch and Taylor (2000), and Pintzuk (2002), they argue that apparent variation in the syntax of English, over the course of its history, is best understood as grammar competition—that is, speakers are exhibiting a type of bidialectalism, wherein they simultaneously use different (but obviously very similar) grammars in ways which result in the attested variation.[7] This is a clear instance of historical facts being marshaled to refine synchronic models of grammar, along the lines envisioned by Kiparsky in the quote above. In this case, they "complicate" the synchronic model by providing evidence that multiple grammars can be instantiated in a single individual. A comparable case of historical evidence being used to refine models of the syntactic structure of grammar can be found in the Transparency Principle of Lightfoot (1979), a proposed synchronic constraint invoked to account for aspects of syntactic change.

Structural approaches to universals and change in the generative tradition often put a strong emphasis on the connection between acquisition on the one hand and universals and change on the other. Albright's contribution to the present work is a clear example of this approach. He uses a particular, well-defined model of acquisition to account for the direction of analogical change in paradigms and situates the locus of change within the acquisition process. Something similar can be seen in Lightfoot (1991), which develops a model of the acquisition of syntax consistent with the Principles and Parameters approach and which is also consistent with observed patterns of historical change. Such work need not necessarily make distinct predictions from models of historical change which are agnostic as to who the agents of change are. But where it can often crucially differ is the emphasis it places on how a change from one grammatical structure to another across (idealized) generations may be triggered by the linguistic forms a language learner happens to be exposed to.[8]

The contribution in the present volume most readily associated with the structural approach to universals and change is that of Kiparsky. His chapter is, at least partially, a response to historically oriented approaches (to be discussed in section 1.5) which have argued against the general-validity structural approaches. He suggests that apparent conflict between these approaches, perhaps, does not reside in how to interpret the linguistic facts but, rather, how to understand the term *universal*. Accordingly, he offers an operationalized definition of the term and then examines whether various putative "universals" are true universals or simply typological generalizations. While he takes the former to result from structural properties of grammar,

[7] A useful summary of approaches making use of grammar competition can be found in Pintzuk (2003: 518–519).

[8] See, for example, Lightfoot (2003: 107), who writes, "the only way a different grammar may grow in a different child is when that child is exposed to significantly different primary data."

he believes that the latter may, in some cases, be best explained as epiphenomena of recurrent patterns of historical change. This is clearly an interesting result, in the present context, since it points the way to a research program in which structural and historical analyses of typological patterns are seen not as antagonistic but, rather, as complementary.

Whitman's analysis of cross-linguistic word-order patterns takes a very similar approach to that of Kiparsky. He classifies word-order universals into three types, *cross-categorial*, *hierarchical*, and *derivational*. The most famous word-order "universals", those establishing a correlation between the order of heads of different syntactic categories and their complements (for example, between verb-object and adposition-object), are classified as cross-categorial. He argues that since these patterns are not absolute but statistical, they should be explained diachronically instead of being analyzed as predictable from the nature of synchronic syntactic structure. However, he further suggests that his other two classes of word-order universals should, in fact, be explained by appealing to the nature of syntactic structure. Thus, like Kiparsky, he views grammatical structure as playing a crucial role in explaining certain attested patterns but also believes that structure should not be invoked to explain all apparent "universals".

Albright's account of analogical change would also seem best classified as structural, since it is grounded in a synchronic grammatical model of paradigm structure and acquisition. It rests on two broad assumptions: (i) that paradigms are organized in speakers' grammars around a single surfacing form which serves as the base for all forms in the paradigm, and (ii) that speakers will choose this single form from the pool of surfacing forms by determining which one allows them to most straightforwardly predict the shape of all the forms in the paradigm. These two proposed principles are understood to manifest themselves during acquisition but are nevertheless taken to be part of the structure of grammar. Therefore, while the kind of data Albright focuses on is purely diachronic, the burden of explaining attested diachronic pathways is placed within synchrony, following the schema in (1b), not (1a). Even though the phenomena that Albright is concerned with may not be the prototypical foundation for a structural explanation, he shows that a sufficiently explicit model of the structure of grammar can go quite far in accounting for them, recalling the point made by Kiparsky (1982: 57), cited above, that progress in historical linguistics may often depend on sharpening synchronic theories.

1.5 HISTORICAL APPROACHES

Historical approaches to universals claim that a particular universal can be understood as a predictable result of attested patterns of language change. This approach was schematized in (1a) where the locus of universal patterns was depicted as

deriving from the ways in which grammars transition between different states. In such a model, even robust synchronic universals may be understood as epiphenomena of similar processes of change applying in converging ways across many languages.

To take an example of this type of approach from Greenberg, one way to explain the synchronically observed pattern that all languages with nasal vowels also have oral vowels is to invoke a historical generalization that "nasal vowels come from oral vowels, and not vice versa (Greenberg 1978b: 51)." Greenberg (1978a: 71) schematizes one common pathway for the development of nasal vowels along the lines of (2).

(2) VN → ṼN → Ṽ

The pathway in (2) views the rise of phonemic nasalization as the result of a sound change producing allophonic nasalization of an originally oral vowel before a nasal consonant followed by a second sound change where the nasal consonant is lost. If we assume that this is the primary mechanism through which nasal vowels develop, we immediately have an explanation as to why languages with nasal vowels also always have oral vowels—the presence of oral vowels in a language is a prerequisite for the development of nasal vowels. Such an explanation makes no appeal to the structure of grammar, only possible directions of change are important, which is why it is labeled *historical* here.

Historical explanations for language universals have a long pedigree, with origins going back to at least the neogrammarians. Paul (1880/1886), for example, explicitly argues against anything but a historical approach to the study of language—and, presumably, therefore, would exclude a synchronic approach to universals entirely. The neogrammarians are generally associated with word-level change (e.g., sound change and analogy), which has meant their work has had more influence on phonology and morphology than other areas of linguistics. However, Delbrück (1880/1974), also in the neogrammarian tradition, employs a similar approach with respect to syntax.

The work of Baudouin de Courtenay and Kruszewski of the Kazan School also shows a tendency for historical explanation. However, it must be readily acknowledged that, as important figures in the development of structural approaches to synchronic analysis, they did work that is properly categorized as simultaneously embracing historical and structural approaches (for general discussion, see Anderson 1985: 56–82). Their position can be aptly summarized with the following quotation: "The mechanism of a language (its structure and composition) at any given time is the result of all its preceding history and development, and each synchronic state determines in turn its further development (Baudouin de Courtenay 1871/1972: 63)."

Despite their historical "head start", historical approaches to universals became significantly less prominent in the twentieth century as the synchronic study of

grammar grew to become a major focus of linguistic theory, first under the influence of de Saussure and, later, under the influence of the generativists. However, even for linguists of the generative tradition, historical explanations were sometimes taken to be the best way to account for certain kinds of widespread grammatical phenomena (if not true universals) which, for one reason or another, resisted straightforward explanation via synchronic models of grammar—and, as discussed in section 1.4, this is also true of some of the structurally oriented contributions in this volume.

One example of a phenomenon discussed in this regard is split ergativity. There are a number of "universals" which can be stated about the nature of split ergative systems. Anderson (1977: 329–330), for example, states that "languages may have ergative marking in perfect (or past) tenses and accusative marking in imperfective (or non-past) tenses, but not vice versa." He suggests that the source of the explanation for this generalization is "to be found in the principles by which perfect tenses are created (330)." (See also Anderson 1989: 343–349.) This clearly is an example of historical explanation of a typological pattern, even if the relevant "universal" is a relatively narrow one.[9] Hyman (1977) is another example of a linguist associated with the generative tradition appealing to diachrony to explain certain phenomena, in this particular case, phonological ones.

However, although the broadest trend of twentieth-century linguistics may have involved a movement away from historical explanations for universals, a number of linguists maintained such approaches. Almost certainly the most significant figure espousing historical explanations was Joseph Greenberg. In a number of works, including Greenberg (1966b) and Greenberg (1978a), he argues that certain basic mechanisms of change are universal to language and that many apparent synchronic universals are the result of common paths of change being instantiated across many languages. Within this book, the chapter by Bybee is most closely aligned with the Greenbergian view—though quite similar views can also be found in the chapters by Blevins and Garrett.

While not always explicitly tied to the pursuit of explanations for universals, the study of grammaticalization (see, e.g., Heine et al. 1991 and Hopper and Traugott 1993 for an overview) should also be mentioned here since work in this area has been used to support the historical approach to universals. The most comprehensive work combining both strands of research is probably Bybee et al. (1994), an examination of the historical development of tense, mood, and aspect marking based on an extensive cross-linguistic survey. One conclusion of this study is that, with respect to the semantics of grammatical morphemes, universal patterns are far better explained through diachronic models than through synchronic ones (Bybee et al. 1994: 281).

[9] Kiparsky's contribution contains a detailed criticism of a comparable historical approach to split ergativity, found in Garrett (1990a).

As Kiparsky mentions in his contribution, historical explanations for universals seem to be "recently regaining popularity", citing work like Bybee (1988a), Garrett (1990a), and Blevins (2004a), as well as Aristar's (1991) study of Greenbergian word-order correlations—not surprisingly, three of the authors he gives are represented in this book since their research speaks so directly to the relationship between diachrony and language universals. Of the above works, Blevins (2004a) is notable for thoroughly codifying a historical approach to phonological universals, which she names Evolutionary Phonology. Her basic approach has been adopted by a number of other phonologists in recent work, including Guion (1996), Kavitskaya (2002), Barnes (2002), and Yu (2003), among others. However, historical explanations for universals in phonology are not limited to such an "evolutionary" approach. Bybee (2001), for example, offers a rather different—though not necessarily contradictory—diachronic model of phonological development from that found in Blevins (2004).

There is, however, an interesting contrast between approaches like that of Bybee (2001), on the one hand, and Blevins (2004a), on the other. While some aspects of their methodology are quite similar, their guiding principles appear to be quite distinct. In the beginning of her contribution to this volume, Bybee writes, "The true universals are the mechanisms of change that create the diachronic paths." Blevins, on the other hand, gives as the central premise of her approach that "[p]rincipled diachronic explanations for sound patterns have priority over competing synchronic explanations unless independent evidence demonstrates, beyond reasonable doubt, that a synchronic account is warranted."

Both approaches have a common diachronic "bias". However, while Blevins explicitly admits the possibility that a synchronic account may be necessary for universal patterns in some cases, Bybee is clearly skeptical about this. Furthermore, Blevins (2004a) generally employs a more or less traditional model of sound change which treats it as a transition from one discrete state to another—a relatively comfortable conceptualization from the point of view of generative/structuralist phonology. For Bybee, however, change is conceptualized much less discretely: "[T]he cumulative effect of [the application of the mechanisms of change] over multiple usage events creates grammar."[10]

In the present volume, both Blevins and Bybee offer historical explanations for universal patterns in line with their previous work. Blevins approaches consonant epenthesis within the framework of Evolutionary Phonology. Bybee develops a historical, usage-based account of a well-attested pattern of phonological alternation known as Structure Preservation, wherein morphologically conditioned phonological

[10] An additional feature of Bybee's paper, worth mentioning in the present context, is the fact that she concludes with, "structural properties . . . arise as language is used and find their explanations in the nature of the categorization and processing capacities of the human brain." While the argument of her paper, therefore, focuses primarily on linguistic mechanisms of change, she clearly views those mechanisms as connected to broader aspects of human cognition, adding a dimension of external explanation, of the sort discussed in section 1.6, to her account.

alternations show an overwhelming tendency to be restricted to contrastive features in a language's phonological system (and, therefore, tend to involve changes from one phoneme to another). This is opposed to, for example, phrasally conditioned phonological alternations, which typically involve only non-contrastive features (and, therefore, involve alternations between allophones of a single phoneme).

While the label given this phenomenon implies it should be more amenable to a structural than a historical account, Bybee argues that a historical one is to be preferred because it can explain both the generally observed pattern as well as known exceptions to it. The very frequently encountered idea that patterns which are not exceptionless might, in general, be better explained historically than structurally is, in fact, a common theme in this volume and is discussed as well by Kiparsky, Kuteva and Heine, and Whitman.

Garrett's contribution also offers a historical explanation for a universal pattern—in his case for the phenomenon known as paradigm uniformity whereby analogical change tends to affect paradigms whose base forms alternate in ways that reduce those alternations. Garrett explicitly argues against any structural property of grammar favoring uniformity over non-uniformity, arguing instead that apparent uniformity effects are epiphenomena of other mechanisms of change.

As discussed in section 1.2, Harris takes on the problem of explaining typologically unusual structures, trying to develop a model which can predict that such rara can exist, on the one hand, but that they would be attested only infrequently, on the other. Her explanation is grounded in the idea that some grammatical patterns are rare simply because the historical chain of events required for them to develop involves a large number of independent changes that would not be expected to "come together" in the right way very often due to the laws of probability. A rare pattern, therefore, may not be rare because there is anything "structurally" wrong with it. Rather, the odds may simply be stacked against its ever arising in the first place.

Kuteva and Heine also offer a historical explanation for certain grammatical patterns, specifically arguing that some instances of apparent grammatical "irregularities" can be straightforwardly explained if we understand the nature of the grammaticalization processes that produced them. However, another dimension to their explanation involves the role of language contact—a factor external to grammar. Accounts for universals invoking such factors will be taken up in the next section.

1.6 EXTERNAL APPROACHES

In addition to structural and historical approaches to the explanation of universals, it seems worthwhile to recognize a third possibility: that the locus of explanation lies in principles not specific to language but, rather, in ones external to it. Of course, we

must readily recognize that an explicitly invoked "linguistic" principle may itself have an ultimate explanation which would be non-linguistic. This issue was briefly taken up at the end of section 1.3. Nevertheless, one can distinguish between approaches which *formulate* their principles as being specific to language as opposed to those whose principles are explicitly understood to be more widely applicable, either to communication in general or more broadly to the human condition. I discuss some relevant approaches of this latter kind here.

Ohala's (1993) model of sound change is a good example of an approach making use of such external principles. It views sound change as resulting from a listener's misanalysis of the phonological representation that a speaker intended for a given utterance, where the range of predicted misanalyses is connected to well-attested types of variation found in the phonetic signal. Critically, Ohala grounds his theory of phonetic variation in a physiological model of speech production. To the extent that human physiology should be viewed as "outside" of grammar, Ohala's model would seem to constitute, at least partially, an external explanation for the relationship between universals and change. Of course, as with some of the contributions in this volume, Ohala's model of sound change does not solely fit into just one explanatory mold. The idea, for example, that a crucial step in language change also involves the *phonological* analysis of the phonetic signal of an utterance means that there is a structural dimension to his explanation as well.

Importantly, the distinction between an external explanation and a structural or historical one can be highly sensitive to theoretical interpretation. Hayes and Steriade (2004), for example, make use of some of the insights of the work of Ohala just discussed. However, they argue not for a model based on "misapprehension" (Ohala 1990: 244) on the part of the listener but, instead, for one where the speaker actually has a "partial understanding of the physical conditions under which speech is produced and perceived" and where this knowledge is actually part of grammar (Hayes and Steriade 2004: 1). They further propose that such grammatical knowledge can drive historical change (Hayes and Steriade 2004: 27). Under their model, some of the explanatory principles treated as external within Ohala's model would be labeled structural within the classification developed here.

External principles have been invoked to account for linguistic phenomena in a number of domains. Bybee and Moder (1983: 267), for example, propose that linguistic objects can be classified on the basis of their phonological form in a way that is analogous to the categorization of natural and cultural objects and suggest that this indicates that some of the principles governing linguistic behavior may, in fact, be more general in nature. Haiman (1983: 816) makes comparable claims for the relationship between morphosyntactic categories and conceptual categories. Similarly, Sweetser (1990: 23–48) makes use of an externally oriented principle of metaphorical extension, taken to be rooted in broad aspects of human cognition, which both delimits possible synchronic metaphorical uses of certain words and also guides how the semantics of words can change diachronically.

More generally, Haspelmath (1999a) has discussed the possibility of explaining the existence of supposed structural grammatical constraints of various kinds, proposed within the framework of Optimality Theory, by appealing to external functional considerations. Specifically, he argues that speakers will choose to use "good" variants of linguistic forms instead of "bad" ones and that these adaptive choices may become entrenched as constraints on grammars. The criteria distinguishing good variants from bad ones, in his view, are not specifically linguistic. For example, he proposes that the fact that many languages allow topic arguments to be unexpressed is connected to a general human proclivity to "save production energy" (Haspelmath 1999: 197).

Haspelmath's arguments raise an important issue with respect to external explanations of universals: the explanatory principle need not be "confined" to either synchrony or diachrony. Synchronic pressures (for example, a functional pressure to save energy) may cause languages to change in particular functional directions (for example, a strong grammatical constraint against expressing topical arguments), resulting in universal patterns. This sort of external approach, which sees synchronic external pressures as the driving forces behind convergent diachronic changes, of course, speaks directly to the theme of the present volume. More general work along these lines includes Keller (1994) and Durie (1999).

In addition to external approaches invoking functional or cognitive constraints, sociolinguistic constraints have also been proposed as playing a role in shaping universal patterns. Labov (2001: 511–518), for example, contains relevant discussion of some possible sociolinguistic principles which relate language change to social perceptions of language. One example of such a principle of social perception is the *Golden Age Principle*: At some time in the past, language was in a state of perfection (Labov 2001: 514).[11] This principle is intended to explain, among other things, why older generations do not typically adopt speech norms of younger generations. Such a sociolinguistic attitude hardly seems to belong to anything like a universal grammar but has clear implications for both synchrony and diachrony and would, therefore, appear to be an externally oriented account of a cross-linguistic grammatical generalization.

Similar work within sociolinguistics includes Trudgill (1989, 1996) and McWhorter (1998), which argue that there may be a connection between a language's sociohistorical profile and its typology (succinctly exemplified in the title of McWhorter 2001, "The world's simplest grammars are creole grammars").

Related to this, of course, is work on the relationship between culture and linguistic patterns (see, for example, Enfield 2002 or Evans 2003 for an overview of relevant research). Much of the work in this area tries to show that the presence of a particular cultural trait in a community may explain the presence of some fairly specific

[11] The Golden Age Principle, of course, is not a true principle of historical linguistics. Rather, it is taken to be held by speakers in a way which informs their attitudes towards language change.

grammatical patterns in that community's language (see, for example, the discussion in Evans (2003: 23–27) on the rise of grammatical encoding of kinship relations in certain Australian languages). However, Enfield (2002: 20) raises the idea that even some apparent grammatical universals may actually be the result of cultural universals, citing the animacy hierarchy as a possible example (see Kiparsky's contribution to this volume for further discussion of this phenomenon—though from a much different perspective).

In the present volume, Haspelmath's paper most explicitly makes use of an external explanation for grammatical universals, in a way similar to Haspelmath (1999a) discussed above. Specifically, he invokes a *functional* principle of economy which causes humans to behave "purposefully and rationally in selecting from available variants and in creating new variants". Over time, this process of selection is taken to lead to the creation of more economical language structures. Crucially, he does not attribute this trend for economy to a grammatical constraint but, rather, connects it more generally to human behavior.[12]

Hopper also invokes an external explanation in his paper, though the principles he employs are of a somewhat different type than Haspelmath's. Specifically, he argues that the exigencies of discourse have played a crucial role in the development of an English construction that he labels *take NP and*. Discourse needs are not part of grammar proper, but rather the communicative situation, making his an external account.[13] In addition, Hopper argues that examining this construction in English can give us insight into similar constructions found in other languages, thus suggesting that not only this one English construction, but a cross-linguistically identifiable set of constructions can be explained by appealing to how discourse requirements can shape grammar.

As discussed above, Kuteva and Heine's contribution contains an element of historical explanation within a grammaticalization framework. However, at the same time, they argue that coming to a full understanding of a particular grammaticalization scenario may require acknowledging the role language contact can play in fostering such a change. In particular, they propose that they are able to improve the predictive power of their grammaticalization model by suggesting that the history of a given language may instantiate a grammaticalization pathway not simply because it was "available" but because its contact relationships—a factor external to grammar—pressured it to develop in that direction.

[12] As with the above example of Ohala's model, there exist analyses very similar to Haspelmath's which do treat his proposed externally oriented principle as specifically part of synchronic grammar. Aissen (2003), for example, who, like Haspelmath, discusses morphosyntactic asymmetries, directly incorporates economy "constraints" into an Optimality Theory model of grammar, without connecting them to human behavior generally, thus indicating she intended her principles to be interpreted structurally.

[13] This is not to say that particular discourse strategies are necessarily outside of "grammar". Only the communicative imperatives shaping a given stretch of discourse are what are taken to be external to grammar here.

1.7 CONCLUSION

Categories like *structural*, *historical*, or *external* are, of course, primarily heuristic in nature. And, in fact, one of the more important conclusions that comes out of this volume is the extent to which a full explanation for the relationship between language universals and language change requires integrating different approaches. Kiparsky, for example, writes from a structuralist perspective but also quite clearly does not see structural explanations as giving us all the answers (see Newmeyer 2005 for a similar view). Bybee, though explicitly taking on a diachronic-functional perspective, nevertheless examines a generalization which was uncovered using structuralist methodology. And, while Haspelmath's paper sees the ultimate explanation for certain phenomena as being external to grammar, he still views the mechanisms through which they become expressed in language as being historical in nature.

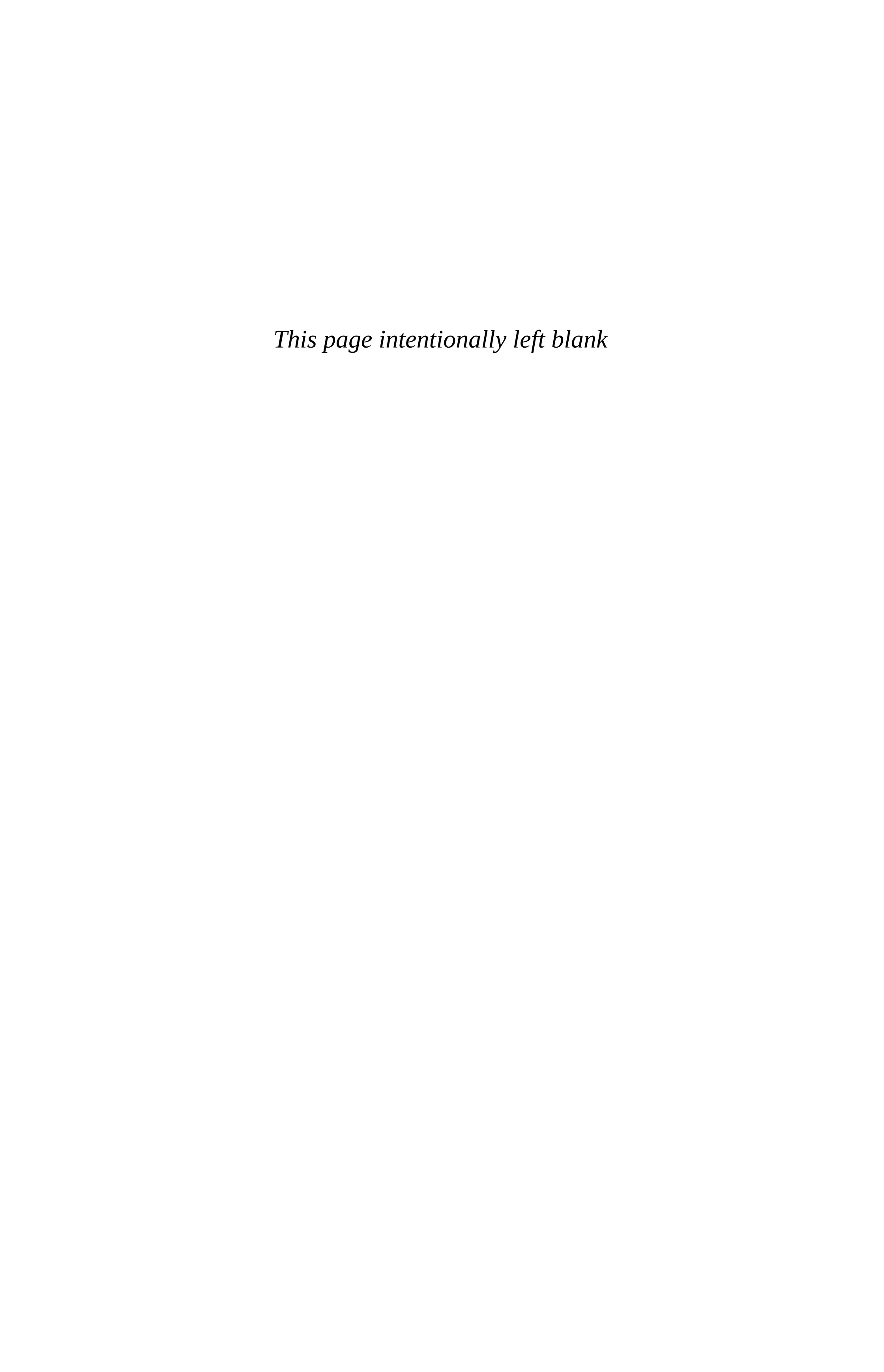
Whereas purely synchronic studies often allow a particular linguist to take on only a "formal"(= structural) or a "functional"(= diachronic/external) perspective (see Newmeyer 1998 for an overview of this issue), it is clearly much more difficult to do so when looking at the relationship between universals and change. On the one hand, the generalizations described by everyone's "universals" need to be explained somehow, even if there is disagreement about how universal they might be. On the other hand, whenever the role of diachrony is introduced into "explanation", language use is generally involved at some level because it is so often implicated in language change—whether its role is limited to the accidental skewings of language input during language acquisition or it is viewed as relevant over the entire lifetime of an individual.

The study of the topic of universals and change is, therefore, not solely interesting in and of itself but is also an interesting arena in which competing linguistic methodologies can be readily compared, allowing us to see what kinds of approaches are well-suited to dealing with what kinds of problems.

This page intentionally left blank

PART I

Universals and Change: General Perspectives



This page intentionally left blank

2

Universals Constrain Change; Change Results in Typological Generalizations

Paul Kiparsky
Stanford University

2.1 THE RELATION BETWEEN SYNCHRONY AND DIACHRONY

2.1.1 Structure explains change

If language change is constrained by grammatical structure, then synchronic assumptions have diachronic consequences. Theories of grammar can then in principle contribute to explaining properties of change, or conversely be falsified by historical evidence. This has been the main stimulus for incorporating historical linguistics into generative theorizing.

A widely shared assumption is that certain mutations occur in the transmission of language. Specifically, they occur when aspects of grammars based on incomplete data, or outputs of such grammars, can be retained from earlier stages of acquisition and become incorporated into the final system. This notion of "imperfect learning" has provided the basis for one approach to analogical change, and, coupled with the theory of Lexical Phonology, provides a solution to the problematic type of phonological change known as lexical diffusion (Kiparsky 1995). It is also commonly assumed in investigations of syntactic change. The theory of acquisition thereby becomes a crucial link between synchronic and diachronic linguistics.

The specific implementation of this approach will depend on the model of grammatical description that is adopted. Syntactic change, for example, has been treated as parameter-resetting (Lightfoot 1991), as grammar replacement (Kroch 1989b), and as constraint reranking (Optimality Theory, recently especially in its stochastic variety, Jäger and Rosenbach 2003; Clark 2004). Each comes with different commitments about the causes and mechanisms of change and about how change is related to synchronic variation. Specific theories of syntax make further predictions about

co-variation between different aspects of grammar, notably between morphology and syntax. For example, on some versions of syntax, rich inflectional morphology entails a highly ramified structure of functional categories to which categories move to check their features, predicting that loss of verb agreement entails loss of V-to-I movement (e.g., Vikner 1995). In a different framework, I have argued that structural position and inflectional morphology are alternative argument licensers, from which I derive, among other consequences, the Sapir/Jespersen generalization that loss of inflectional morphology entails fixed order of direct nominal arguments (Kiparsky 1997).

The leading idea behind this work, that properties of language change might be explained by the way language is acquired and structured in the mind, is of course by no means original to generative grammar. The neogrammarians, for example, had recognized the pervasive role of analogy as a regularizing force in change as a manifestation of the mechanism that underlies the normal acquisition and creative use of language. The structuralists on their part sought to derive the empirical generalizations about language change discovered by the neogrammarians from the design features of language. Indeed, the very origins of structuralism lie precisely in this attempt to ground historical linguistics in a new understanding of the language faculty.

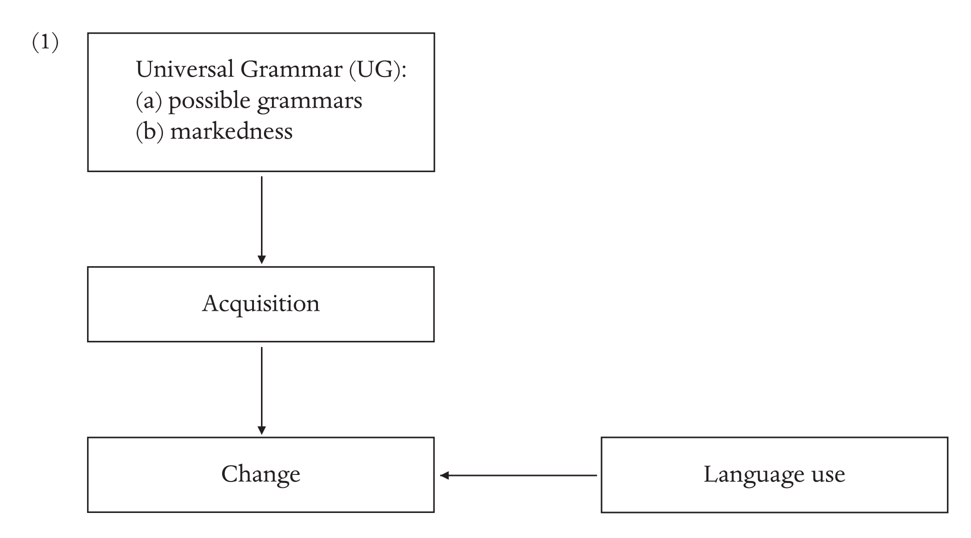
One explanatory connection between linguistic change and the organization of language that emerged in this first round of structuralist theorizing was that language is a network of syntagmatic and paradigmatic relations which define the tracks of potential analogical changes. First articulated by the neogrammarian Hermann Paul, its larger theoretical consequences were worked out by linguists such as as Kruszewski, Baudouin de Courtenay, de Saussure, and later by Jakobson.

A second major idea that emerged at this time was that the regularity and exceptionlessness of sound change discovered by the neogrammarians is based on the independence of phonology from morphology, syntax, and semantics. In Saussure's formulation, the reason sound change is regular is that the link between expression and meaning constituting the sign is arbitrary. Bloomfield's version of the explanation is based on the notion of separation of levels, and in particular on the premise that the phonological and morphological organization of language are independent:

> Theoretically, we can understand the regular change of phonemes if we suppose that language consists of two layers of habit. One layer is phonemic: the speakers have certain habits of voicing, tongue-movement, and so on. These habits make up the phonetic system of the language. The other layer consists of formal-semantic habits: the speakers habitually utter certain combinations of phonemes in response to certain types of stimuli, and respond appropriately when they hear these same combinations. These habits make up the grammar and lexicon of the language. (Bloomfield 1933: 364–365)

So, the structuralist/generative program for historical linguistics during most of the last century looked something like what is given in (1) (read the arrows as "explains" or "constrains").

(1)



For the structuralists, UG (they never called it that, of course) tended to be very simple, and in principle derivable from a few quite general relations. Bloomfield and de Saussure take distinctness (or contrast) as the basic relation in language, and connect it to the property of arbitrariness (Bloomfield 1933: 144).[1] When seriously explored in descriptive practice by post-Bloomfieldian structuralists, this minimalistic program based on distinctness turned out to be problematic, as became clear first in phonology and then in morphology and syntax. Generative grammar therefore ended up positing a rather richly structured language faculty as an innate endowment.

This research program does not foreclose functional explanations for language change. Uncontroversially, functional factors shape language use. The generative program opens up the possibility that they might have become biologized within UG itself, thereby constraining change also via acquisition.[2] For example, we could speculate that evolutionary pressures might have caused the innate learning mechanism to favor grammars that optimize perception, production, and/or stable transmission in certain ways. A language designed in modular fashion, with different levels of

[1] As an example, here is Bloomfield's definition of *part of speech* (Bloomfield 1926: Def. 38): "The maximum word-classes of a language are the *parts of speech* of that language". To apply the definition, you have to go back to the definitions of *maximum, language,* and *word-class.* A *maximum* X is an X which is not part of a larger X (Def. 26). A *language* is the totality of utterances that can be made in a speech community (Def. 4, with Defs. 1 and 3, and Assumption 1). A *word-class* is a *form-class* of *words* (Def. 37). A *form-class* is the set of forms having the same functions (Def. 33). A *form* is the set of vocal features common to same or partly same utterances (Def. 6). *Same* is that which is alike (Def. 5). A *function* is the positions in which a form occurs (Def. 32). A *position* is an ordered unit in a construction (Def. 29). A *construction* is a recurrent same of order (Def. 23, with Ass. 8). A *word* is a minimum free form (Def. 11). A *free* form is a form which may be an utterance (Def. 10). And so on, all the way down.

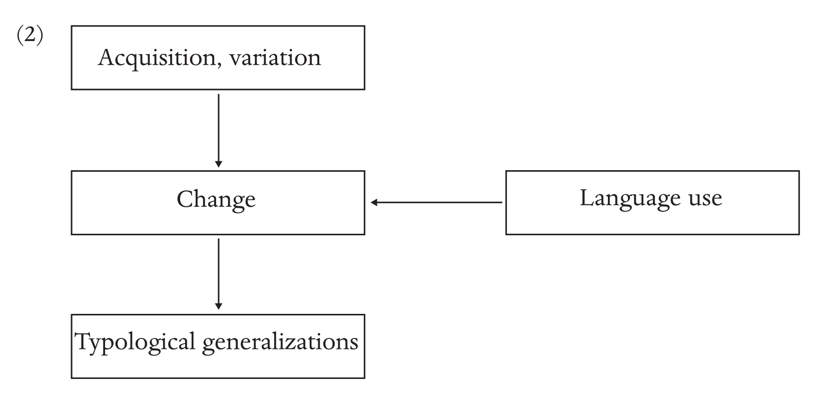
[2] The idea that the language faculty has evolved to cater to the speech production and speech perception systems is obviously true on the physical side of speech. The mammalian ear (including famously that of a chinchilla) is especially good at distinguishing certain kinds of sounds, and the vocal tract of humans—the only mammal with speech—has become biologically adapted to producing sounds rapidly and accurately in just that region of acoustic space.

representation subject to their own constraints, may well be the most efficient for this combination of tasks.

2.1.2 Change explains structure

Seemingly at odds with the paradigm in (1) is an older, pre-structuralist idea which is recently regaining popularity. It views the direction of explanation as going the other way: cross-linguistically recurrent structural patterns in grammar are due to recurrent patterns of language change (Bybee 1988b, this volume; Garrett 1990a, this volume; Aristar 1991; Blevins 2004a, this volume; Kuteva and Heine, this volume).

(2)



In its more conservative forms, this program sees historical explanation as a kind of supplement or corrective to the formal theory of grammar. Most commonly, it is claimed that UG should be seen as a theory of CORE GRAMMAR, and that vicissitudes of language change under some circumstances can produce "marked" or even anomalous structures which fall outside the remit of UG. Such is, for example, a common view of split ergative case systems, which have no straightforward analysis in Government and Binding (GB) and its successor theories (see section 2.3 below). Or, conversely, it is claimed that the formal theory of language should *overgenerate* by allowing for possible types which are unattested simply because they cannot arise—or at least cannot easily arise—through normal processes of change (Harris, this volume).

In its more radical form, the program does not seek to complement the theory of UG but to replace it. It tries to explain away putative universals as by-products of recurrent patterns of language change. As Aristar puts it, "different diachronic processes together conspire to give the effect of synchronic universals" (1991: 5). As a case study he offers an account of the Greenberg word-order universal in (3).

(3) Genitives, relatives, and adjectives usually precede their heads in SOV languages and follow them in VSO languages.

The basis of (3) (according to Aristar's historical reinterpretation) is simply a *historical* relationship between the three categories:

[A]djectives, relatives, and genitives pattern similarly because relatives and genitives potentially have their diachronic source in a binding-anaphor construction . . . and that adjectives pattern similarly to these because adjectivals often have their diachronic source in genitives and relative clause. (Aristar 1991: 26)

The competing synchronic explanation for (3) is of course that genitives, relatives, adjectives, and subjects are specifiers, and that, by cross-categorial harmony (Hawkins 1983) all specifiers will normally either precede or follow their heads.

In this particular case, the synchronic structural explanation based on cross-categorial harmony seems to have the edge over the historical explanation. It correctly extends to adjectives from other sources than genitives and relatives, for which the word-order correlation stated in (3) is just as valid. Moreover, the fact that genitives and adjectives tend to go hand in hand in word-order *change* (as they have in the history of English) indicates that the syntactic relation between them is intrinsic. Furthermore, the parallelism is more fine-grained than common origin could explain, and there are systematic disparities which are subject to additional generalizations (Dryer 1988, 1992). Giorgi and Longobardi (1991: ch. 3) show that at least some of these additional generalizations can be explained within the theory of grammar by independently motivated assumptions about the phrase structure of nominals.

Moreover, historical explanations, once spelled out, often turn out to appeal implicitly to tendencies that are themselves in need of explanation. For example, without a theory of categories and phrase structure, the direction of reanalysis which Aristar takes as a given is just as puzzling as the typological universal it is supposed to ground. For a true explanation we need a theory of phrase structure and grammatical categories. Genitives and relatives are indeed among the diachronic sources for adjectives, but the fact that genitives and relatives are likely candidates for reanalysis as adjectives is itself a puzzle, not something self-explanatory. One plausible answer is based precisely on their common status as nominal specifiers.

2.1.3 The program

The opposing claims of synchronic and historical explanation lead to a research program which integrates historical and synchronic linguistics by demarcating in a principled way the explanatory role of each. When is change the explanans, when is it the explanandum? Answering that question will give a basis for distinguishing true universals from typological generalizations. The issue goes well beyond the simple question how cross-linguistic generalizations originate. It is about the nature of those generalizations themselves. Whatever arises through language change can be lost through language change (unless it gets somehow incorporated into the genome, as mentioned above). Any structural feature that is caused by change is inherently unstable (vulnerable, as Saussure put it). It can be washed out by other changes, or replaced with the opposite feature. Therefore recurrent structural features that are

caused by recurrent patterns of change are TYPOLOGICAL GENERALIZATIONS but not true universals.

By the same token, if a generalization is itself a determinant of historical change, it must be a true intrinsic UNIVERSAL, which is properly the subject matter of UG.

What is a language universal? In the typological literature it is usually taken to be an absolute or implicational generalization which is true for all languages. In some versions of generative grammar, aspects of phrase structure have been proposed as universals in a similar sense. For example, it has been proposed that all languages have the same functional categories organized in the same hierarchical phrase structure at Deep Structure.

A rather different conception of language universals emerged in Jakobson's and Trubetzkoy's work. When Jakobson presented a set of distinctive phonological features as universal, he of course did not mean that each of these features is *universally instantiated*, but rather that each of them is *universally available*. The same is true of the set of distinctive features that he proposed for morphology. Stampe's universal phonological processes (see, e.g., Stampe 1979) have much the same character, except that he was more explicit about the mechanisms by which they could become suppressed within a language.

A kindred view was later elaborated in Optimality Theory (OT). Like Jakobsonian features, Optimality-theoretic constraints are universal but not necessarily visible in all languages. A universal constraint can be violated just in case a more highly ranked universal constraint requires it. Therefore, constraints are violable, but violations are minimal. Language-specific properties are thus essentially reduced to constraint ranking, and in principle every possible constraint ranking defines a possible language. Hence the totality of possible constraint rankings defines the space of possible grammars (factorial typology).

Here I adopt the OT view of universals as constraints which can be violated only by being superseded by some other universal constraint which is more highly ranked in the grammar. If constraints are universal in this sense, it follows that effects of markedness constraints occulted by dominant constraints will become visible in contexts when those higher-ranking constraints are not applicable (THE EMERGENCE OF THE UNMARKED, OR TETU, Prince and Smolensky 1993). For example, even a language that permits complex syllable structure may reveal the latent universal preference for CV in neologisms, reduplication, and other contexts where the shapes of syllables are not specified in the lexicon or dictated by higher-ranking constraints. Such TETU effects, conceptually akin to Stampe's argument for innate processes, constitute powerful evidence for the universality of constraints.

While universal constraints are part of every grammar, there may also be "accidental" generalizations which do not correspond to any constraint or rule of grammar, even though they may be true of the sentences of a language. Positive evidence of the reality of a generalization is its productivity and its role in the system, revealed by interaction with other grammatical generalizations.

The question whether explanations are to be located in synchrony or diachrony is in principle independent of the question whether those explanations are functionally grounded, and, if so, how. A synchronic explanation is based on some feature of language design. This may be grounded in some inherent property of the human mind, either specific to the language faculty, or characterizing cognition in general. Or the explanation may have a direct functional grounding in the requirement that speech should be effectively produced and understood, and/or that languages should be readily learnable.

Summarizing the above discussion, and anticipating what follows, let us posit the tentative criteria in (4).

(4)	*Universals*	*Typological generalizations*
	No exceptions	Allow exceptions
	Convergence	Single source
	TETU effects	No TETU effects
	Manifested spontaneously in child language	Not manifested in child language
	Pathways of change	Inert
	Part of the grammar	Not necessarily part of the grammar

I will now apply these criteria to a number of proposed typological generalizations and candidate universals. The material is drawn from both phonology and syntax; many of the cases involve a scale or hierarchy which defines the parameter of a rule or constraint. The results show, I think, that the criteria in (4) converge fairly neatly to sort out the true universals, in the above sense, from the typological generalizations.

2.2 MORPHOLOGY AND BINDING PROPERTIES OF REFLEXIVES

2.2.1 Simple and complex reflexives

A case where I think diachrony convincingly explains a set of typological generalizations has to do with the relation between the morphological properties of anaphors and their binding behavior. Here I will be drawing on the typology of anaphora proposed in Kiparsky (2002) (and see Gast 2006).

Reflexive pronouns are of two main morphological types, SIMPLE and COMPLEX. Simple reflexives are typically monomorphemic elements, such as French *se*, German *sich*, Russian *sebja*. Complex reflexives are of two types: (1) the HEAD-type, which consists of a possessive pronoun combined with an inalienably possessed noun, typically 'head' or 'body', and (2) the SELF-type, which consists of a reflexive or pronominal combined with an adverb that means 'self' (German *selbst*, Swedish *själv*, French *même*, Italian *stesso*, Russian *sam*).

A typological generalization discovered by Faltz (1977) and theoretically explored by Pica (1987) (sometimes called "Pica's generalization") states that complex reflexives typically differ from simple reflexives as in (5).

(5) a. They allow object antecedents.
b. They must be bound locally within the same clause.
c. They typically lack possessive forms.

Synchronic explanations have appealed to the idea that simple and complex reflexives have different syntactic structures, which cause their different behavior. The synchronic explanation for (5) that has been proposed in the literature is that complex reflexives are maximal projections (syntactically complete phrases), whereas simple reflexives are heads (essentially, words). The idea is that simple reflexives "cliticize" to INFL (or some other functional head of the clause) at Surface Structure or at Logical Form (LF). When raised to this position, they are C-commanded only by subjects (Pica 1987; cf. Katada 1991; Hestvik 1992). Long-distance binding takes place by successive cyclic movement to higher positions. Problems with this account include that the posited LF movement would violate both the Coordinate Structure Constraint and the Empty Category Principle (ECP); it is also not clear how to get long-distance binding of reflexives inside maximal projections.

The alternative historical explanation is that complex reflexives arise as anti-obviation strategies. A universal principle of *Coargument Disjoint Reference* (CDR) requires that coarguments (arguments of the same predicate) cannot overlap in reference, unless they are specially marked as exempt from it (Kiparsky 2002). Pronouns like *him* are not so specified, and neither are ordinary simple reflexives. Hence such elements fall under the constraint in (6).

(6) *CDR:* A pronoun cannot overlap in reference with a coargument
a. John hates him. (there must be two people involved)
b. Each of the men hate him. ('he' isn't one of 'the men')

CDR applies not only to referring expressions (nominal and pronominal elements) but also to anaphors, unless they are specially marked as exempt from CDR. (The property of being subject to CDR is referred to as *obviation* in the grammatical literature.)

The two types of complex reflexives, the HEAD-type and the SELF-type, represent precisely the two ways in which a pronoun (whether pronominal or anaphor) subject to CDR can be marked so as to escape it. Head-type complex reflexives defeat this constraint by putting the pronoun into a non-coargument position. Self-type complex reflexives defeat it by marking the pronoun as exempt from CDR (by adding an element which asserts identity between the pronoun and a contextually determined element). The distribution of complex reflexives is restricted to environments where CDR needs to be defeated. The properties of complex reflexives in (5) then follow straightforwardly.

(7) a. Complex reflexives allow object antecedents because they are not subject to CDR.
b. They are bound within the same clause because long-distance antecedents are not coarguments of them.
c. They typically lack possessive forms because a possessor is not a coargument of its possessum's coarguments.

If this explanation is correct, then (5) is not a linguistic universal and should NOT be expressed in the synchronic theory of grammar. This might be a welcome conclusion, because a principled connection between the shape of an anaphoric expression and its binding properties has proved elusive so far. At least on lexicalist assumptions, the syntax has no access to the morphological composition of words. The putative correlation of the morphology with the categorical distinction between heads and maximal projections is stipulative, and empirically suspect besides: typically the distribution of morphologically complex reflexives like *himself* is the same as that of the simple pronouns they contain (such as *him*), so they are probably not maximal projections (Toivonen 2001).

Therefore, in terms of our criteria in (4), what we have here is not a true universal, but a typological generalization with a historical explanation, as discussed in (8).

(8) a. The generalization has arbitrary exceptions, i.e., exceptions not motivated by more highly ranked constraints (Huang 2000: 96).
b. All complex reflexives seem to arise in the same way, by the route described above.
c. There is no acquisition evidence which would show that learners access it.
d. There is no historical evidence which would show that it is analogically generalized.
e. The generalization is probably not structurally encodable.

2.2.2 Nominative anaphors

Another typological generalization is given in (9).

(9) There are no nominative anaphors.

This is trivially true in languages in which anaphors must be locally bound to a nominative subject. But it is not obvious why it would hold even in languages which allow long-distance binding, or in languages in which nominative case may be assigned to objects, such as Icelandic (see, for example, Maling 1984).

(10) Icelandic
a. *Honum finnst (sjálfur) sig (vera) veikur
Him.DAT finds self REFL.NOM (be) sick.NOM
'He considers himself (to be) sick' (no reflexive nominative object)
b. Hann sagði að sig vantaði hæfileika
He said that self.ACC lacked ability.NOM
'He said that he lacked ability' (reflexive accusative subject)

As we'll see, the generalization (9) is actually not true, but it is still a pretty robust tendency, so some explanation is called for. Synchronic explanations that have been proposed for (9) include LF movement subject to the ECP, and several agreement-based hypotheses.

The ECP explanation, due to Chomsky (1986), posits that anaphors move at LF to INFL (i.e., an "inflectional" projection in the syntax), leaving a trace; in subject position the trace would not be properly governed. This does not really account for nominative objects, or for pronominal chains.

Rizzi (1990) and Woolford (1999) proposed instead that nominatives agree with AGR (i.e., an agreement element), which is pronominal. If the nominative were an anaphor, the result would be a chain which would have to be both locally bound and locally free. However, this still won't work for nominative objects.

Everaert (2001) developed a minimalist version of the agreement story, according to which V's uninterpretable ϕ-features must be checked against an agreeing N's interpretable features. Nominative anaphors are excluded if they are not fully specified for some ϕ-feature that must be licensed on the V. For example, Icelandic *sig* is unspecified for number, so it can't check the V's number feature. On the other hand, Georgian and Marathi nominative reflexives are acceptable because they agree.

(11) Georgian (Harris 1981; Everaert 2001)

a. Vano-m daurc̣muna tavis-i tav-i
Vano-ERG he.convince.him.AOR self.GEN-NOM self-NOM
'Vano convinced himself' (reflexive nominative agreeing object)

b. Gela-s turme daurc̣munebia tavis-i tav-i
Gela-DAT apparently he.convince.him.EV self.GEN-NOM self-NOM
'Gela has apparently convinced himself'

c. tavis-ma tav-ma Ø-xsn-a president-i
self.GEN-ERG self-ERG he-saved-him president-NOM
'it was the president who saved himself (no one else did it)' (reflexive ergative agreeing subject)

(12) Marathi
Jane-ne$_i$ John-laa$_j$ kaḷavle ki aapaṇ$_{i,*j}$ turangaat aahot
Jane-ERG John-ACC informed.3SG that self.NOM prison.LOC was.3SG
'Jane$_i$ informed John$_j$ that self$_{i,*j}$ was in prison' (reflexive nominative agreeing subject)

Everaert's solution comes a lot closer, but it still runs into empirical problems. The correlation it predicts breaks down in both directions. Swedish has an unspecified reflexive *(sig)* and no V agreement, yet still lacks nominative anaphors. Choctaw has an unspecified reflexive and rich V agreement, yet does have nominative anaphors (Broadwell 1988).

The historical explanation, admittedly rather unexciting by comparison, starts from the observation that when nominative objects are prohibited and subjects can't be bound outside a finite clause, nominative anaphors are simply impossible. Germanic

and Romance were originally such languages. The morphological gap persisted even after nominative objects and/or long-distance binding arose in some of them, as in Icelandic. Marathi and Georgian, on the other hand, never inherited such a constraint. Marathi seems to have had long-distance binding of anaphors, nominatives included, as long as it has had the reflexive *aapan*. For *aapan* is derived from Sanskrit *ātman* 'soul, self', which there functioned as a reflexive (or, rather, as the equivalent of one) in any case form, including the nominative, as in (13).

(13) khaḍgena śakyate yuddhe sādhv ātmā parirakṣitum
sword.INSTR can.3SG combat.LOC well self.NOM protect
'one can protect oneself well with a sword in combat' ('one's self can be protected well') (*Mbh* 12.160.3)

That is really all that needs to be said. There is simply no synchronic principle at work. The historical explanation covers the data perfectly.

2.3 SPLIT ERGATIVE CASE MARKING

2.3.1 The D-hierarchy

For a case where the evidence seems to point in the opposite direction, let us turn to a characteristic asymmetry of case marking and the so-called animacy hierarchy that determines it.

A case assigned to subjects of transitive verbs but not to subjects of intransitive verbs is called ERGATIVE. SPLIT ERGATIVE systems have ergative case marking under restricted conditions, most commonly depending either on the nature of the NP or on the tense/aspect of the verb. A classical example of an NP split ergative case system is Dyirbal, which has an ergative/nominative opposition in nouns, and a nominative/accusative opposition in pronouns (Dixon 1972). Dyirbal's structural case system is shown in (14), where "A", "O", and "S" denote the subject and object of a transitive verb and the subject of an intransitive verb, respectively.

(14)

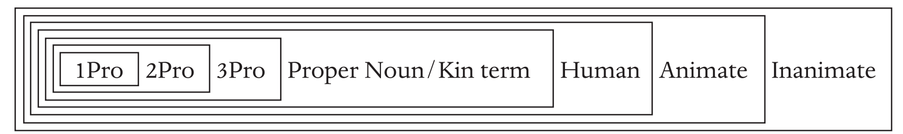
	A	S	O
Nouns	ERG *-nggu, -ru*	NOM -Ø	NOM -Ø
Pronouns	NOM -Ø	NOM -Ø	ACC *-na*

When split ergativity is conditioned by the inherent category of the NP, the cases tend to be distributed according to the hierarchy in (15),[3] which I'll refer to as the

[3] The hierarchy was extensively discussed by Kenneth Hale in lectures at MIT in the late 1960s; see Hale (1973b). Silverstein (1976) and Dixon (1979) documented its application to ergative case systems.

D-hierarchy, since I shall propose to relate it to definiteness and other features of the determiner system, and the more usual term "animacy hierarchy" is misleading.[4]

(15) The D-hierarchy



Ergative is found in nominals on the right end (the "low" end) up to some cutoff-point on the hierarchy, and accusative in nominals from the left (the "high" end).

In Dyirbal, the two case-marking subsystems divide nominals cleanly into two groups, but in some languages the cutoff-points don't coincide, as in (16).

(16) Djapu (Morphy 1983)

	A	S	O
Non-human	*-thu*		
Human	*-thu*		*-nha*
Pronouns			*-nha*

The distribution of structural case marking in some Australian languages illustrates some of the possible cutoff-points (adapted from Blake 1977, 1987).

(17)

<table>
<tr><th></th><th>Pronouns</th><th>Proper / Kin</th><th>Human</th><th>Animate</th><th>Inanimate</th></tr>
<tr><td>Thargari</td><td colspan="4">Accusative</td><td>Erg</td></tr>
<tr><td>Arabana</td><td colspan="3">Accusative</td><td colspan="2">Ergative</td></tr>
<tr><td>Gumbainggir</td><td colspan="2">Accusative</td><td colspan="3">Ergative</td></tr>
<tr><td>Dyirbal</td><td>Accusative</td><td colspan="4">Ergative</td></tr>
<tr><td>Warlpiri</td><td colspan="5">Ergative</td></tr>
<tr><td rowspan="2">Djapu</td><td></td><td colspan="4">Ergative</td></tr>
<tr><td colspan="3">Accusative</td><td colspan="2"></td></tr>
</table>

[4] Wierzbicka (1981) shows that the hierarchy involves neither "animacy" nor "agentivity", which makes a direct functional explanation implausible. A category related to definiteness, such as individuation or "topic-worthiness", is a more likely candidate, as she points out and as I will also argue below. Let us note here that the hierarchy is actually not always so tidy. One somewhat widespread pattern groups kinship terms with the pronouns. Sometimes "animates" are restricted to higher or intelligent animals, the others patterning with inanimates.

2.3.2 Is split ergativity an epiphenomenon of change?

It is often claimed, by "formal" as well as "functional" linguists, that ergative case marking, and specifically split ergativity of this type, is structurally or functionally unmotivated, but arises through understandable diachronic processes.

(18) a. "NP split-ergative systems in fact have their striking synchronic features as a straightforward consequence of their ordinary diachronic source [instrumental case] ..." (Garrett 1990a: 262)
b. "The split nature of many ergative-absolutive case systems looks like another Rube Goldberg feature of grammars, but we can understand how they might have arisen historically." (Lightfoot 1999: 141)
c. "These restrictions [split ergative case-marking patterns] make little synchronic sense for an active-direct ergative clause in which the agent is the more topical—proximate—argument." (Givón 1994: 33)

Let us consider the arguments for this position, which are laid out in Garrett (1990a). Garrett maintains that the lack of ergative marking on the high end of the D-hierarchy neither has nor needs a synchronic explanation. It is only at an antecedent historical stage that the split case marking is motivated. Garrett assumes that those ergative cases which exhibit such a gap in their distribution are historically derived from instrumental cases, and proposes to explain the gap as an inheritance from this stage. The idea is that the gap is a reflection of the fact that instrumental case is for pragmatic reasons most common with inanimate nouns, for it typically denotes instruments which are normally inanimate. After the instrumental-to-ergative reanalysis, the inherited restriction is transferred to the new ergative case. In its new form, however, the restriction no longer has any synchronic motivation, and therefore tends to get eliminated by analogical spread of ergative case to animate nouns. Pronouns, though, tend to be morphologically so different from nouns that they escape their analogical influence. The typical outcome of this scenario, then, is a noun/pronoun split.

I believe that Garrett's purely historical account for the split ergativity patterns is not viable, for the following reasons.

First, the historical scenario involves an unexplained shift from a *semantic/pragmatic* gap before the reanalysis to an *allomorphic* gap after the reanalysis. Actual instruments are restricted to low-D nominals, simply because they *are* inanimate. (It does not matter for our purposes whether this is a pragmatic restriction or an ontological one.) Garrett assumes that when instrumental case is reanalyzed as ergative case, this restriction is carried over, resulting in split ergativity. The assumption that "the distribution of newly-created linguistic categories parallels that of their immediate antecedents" (Garrett 1990a: 286) is reasonable *per se*. But taken literally, it would imply that after the instrumental-to-ergative reanalysis, high-D nominals would lack ergative case. But, at least in most NP split ergative systems, high-D nominals do *not* lack ergative case; rather, they have ergative/nominative syncretism, hence *ergative nominals with no overt case marking*—a very different thing. The suffixless ergative pronouns have exactly the

same syntax as overtly marked ergative nominals: in particular, they agree with them in case, and are treated as parallel with them in conjoined noun phrases. Such a gap in ergative paradigms is a matter of morphology, not of the distribution of the category of ergative case (and still less a matter of pragmatics, of course).

What is missing from the historical account, then, is the causal link between the putative pragmatically motivated gap in the distribution of the former instrumental case and the zero allomorph that it supposedly leaves behind in the paradigm of the reanalyzed ergative case. In other reanalyses this kind of thing generally does not happen.[5] For example, instrumental case itself typically comes from comitative case, via a chain of development which runs from "in the company of" (*John ate cheese with Mary*) via "accompanied by" (*John ate cheese with wine*) to "by means of" (*John ate cheese with a fork)*.[6] At the outset, inanimate nouns would not be used in the comitative (for "pragmatic" reasons), and yet we never find their instrumental offspring sporting zero allomorphs on inanimate nouns (or lacking instrumental case, for that matter). Similarly, ablative cases denoting source-type relations often originate as local cases with a separative meaning; a local separative case would be restricted to nouns denoting physical objects (for "pragmatic" reasons), and yet we don't find ablatives with zero allomorphs on abstract nouns.

A third example is the reanalysis of possessive suffixes as object markers of definite NPs that has taken place in some languages belonging to the Permic branch of the Finno-Ugric family. Accusative case, which was lost in the nominal inflection of the Permic languages (a state of affairs preserved in Ostyak) was renewed in Komi and Votyak by reanalysis of the third-person possessive suffix as an accusative. Yet we do not find that those nouns which for pragmatic reasons rarely have a possessor (let us say 'sky', 'sun') lack accusative endings (let alone accusative case) in these languages.

The second argument against a historical account of the case-marking pattern is that it is not general enough. It addresses only the distribution of ergative/nominative syncretism. But the same hierarchy also determines the distribution of genitive/nominative syncretism, as in Yukagir, where there is no reason to suspect an instrumental origin for genitive case (Krejnovič 1958: 80, 63 ff.; Nichols 1992: 53).

(19) a. met nime
I house
'my house'
b. Beke ile
Beke deer
'Beke's deer'

[5] Indeed, it might be argued that the very possibility of reanalysis is inconsistent with the transfer of a pragmatic gap motivated by the original meaning. You cannot know about a pragmatic gap without having learned the semantics on which the pragmatic inference depends (in this putative case, the instrumental meaning), but once you have learned that meaning, you will not reanalyze that meaning as another meaning.

[6] An example of this development is Estonian *-ga*, etymologically from *kansa-ssa* 'in the crowd'.

c. ile-n jawul
deer-GEN track
'(the / a) deer's tracks'

In the opposite direction, the hierarchy also determines the distribution of accusative / nominative syncretism already mentioned in (17).

(20) Finnish 'split accusativity': only pronouns get accusative case endings
Koira näh-tiin. Häne-t näh-tiin. (Finnish)
Dog.**NOM** see.PASS.PAST He.**ACC** see.PASS.PAST
'The / a dog was seen.' 'He was seen.'

It also governs dative / accusative syncretism and ergative / dative syncretism, and for the same reason.

Finnish child language and certain dialects reportedly extend accusative marking further along the trajectory of (15), from pronouns to proper names and appellatives, e.g., *Lauri-t*, *isi-t* 'daddy'.

Furthermore, the same hierarchy also governs inverse systems, which are typically found with agreement rather than case, where instrumental to ergative reanalysis cannot be at stake. Therefore, we need a more comprehensive explanation for the different manifestations of the hierarchy than what the instrumental-to-ergative reanalysis scenario by itself can provide.

The third objection to Garrett's proposal is that the historical account is insufficiently general even for the distribution of ergative endings because the phenomenon to be explained has several historical sources. It is not even always true that instrumental case prior to its reanalysis as ergative is restricted to inanimates. More often than not, instrumental case has other functions than expressing instruments, such as marking demoted agents of passives, which are prototypically animate, in fact prototypically human. But such instrumental agents are also reanalyzed as ergative subjects in passive-to-ergative reanalyses, as in Indo-Aryan and Polynesian. And ergative cases that are known to have these origins can also lack an overt ergative case morpheme at the high end of the D-scale; this is the case for first- and second-person pronouns in a number of Indo-Aryan languages, including Marathi, Punjabi, Eastern Rajasthani, Assamese, and Siraiki (= Lahanda, Masica 1991: 252). The examples in (21) are from Siraiki (Bhatia 1993: 181).

(21) a. mãi axbaar ðiṭṭhii
I newspaper.F.SG see.PAST.3.F.SG
'I saw the newspaper'
b. huu ne axbaar ðiṭṭhii
he ERG newspaper.F.SG see.PAST.3.F.SG
'he saw the newspaper'

Since the Indo-Aryan instrumental case from which the ergative marker *ne* is believed to descend (but see Butt 2001) was at no stage restricted to inanimates, the high-D

nominative/ergative syncretism in these languages must have arisen in some other way than by Garrett's scenario. It appears that it arose simply by ergative forms being extended to function as nominatives.[7] Analogous developments can be observed elsewhere. Tibeto-Burman starts with a fully ergative case-marking system, and nominative case spreads in an orderly way down the hierarchy (Bauman 1979).

Nor are high-D unmarked ergatives always descended from former instrumental cases (or, for that matter, from any other case whose exclusion from animates could be motivated semantically or pragmatically). Ergative case can also originate as a generalized oblique case, or dative case. This seems true in the Daghestanian (Northeast Caucasian) languages, and in those Northwest Caucasian languages which have developed a case system. In several of them, pronouns (and sometimes in addition a class of high-D nouns) lack overt endings in the ergative, e.g., Adyghe, as in (22) (Rogava and Kerasheva 1966: 59–95).

(22)		'house'	'I'
	Nominative	unə(-r)	sə
	Ergative	unə-m	sə

In Adyghe and the fairly closely related Kabardian and Ubykh, several of the numerous functions of ergative case (recipient, locative, time) would be puzzling for a reanalyzed instrumental case.

High-D ergative/nominative syncretism seems to have yet another source in some Australian languages. The morphologically unmarked pronominal forms used in S and A function are thought to be ergative in origin, as in Indo-Aryan. Dixon (1980) reconstructs an original three-way nominative/accusative/ergative case opposition for Pama-Nyungan. Most of the daughter languages adopted a disyllabic minimal word requirement, as a result of which the monosyllabic nominative pronouns were either lost, or, in a few languages, augmented with an empty second syllable. In those languages where the nominative was lost, its function was taken over by the ergative. The original three-way case contrast survives in that minority of languages which either rescued the nominative phonologically by adding the extra syllable, or which never adopted the two-syllable word constraint in the first place. For Garrett's theory, this development presents the following puzzle. If the two-syllable word minimum is an innovation, then it must have ousted the original monosyllabic *nouns* as well: so why didn't *their* nominatives get replaced by the ergative forms? Again we have a paradigmatic skewing of ergativity marking along pronoun/noun lines. To make sense of the different manifestations of it we must assume that the paradigmatic skewing has a natural basis, and is not just an epiphenomenon of history.

In sum, the third argument against Garrett's position is that there is not just one path to nominative/ergative syncretism. All diachronic roads lead to the same synchronic Rome, where ergative case lacks a morphological mark in high-D nominals.

[7] This development also took place in Hindi, where however the original ergative forms were subsequently renewed by the addition of the suffix *-ne*.

Far from explaining this syncretism pattern, the various changes themselves require a motivation for the pattern as part of their explanation. The "invisible hand" of historical evolution nudges morphological systems towards certain optimal states, and part of the job of morphological theory is to say what those states are.

A fourth objection to the notion that split ergativity patterns are side effects of the historical change from instrumental case to ergative case is that it predicts the wrong split. Any pragmatic restriction on instrumental case would have to do with animacy, so that ergative case in its pristine, native state should lack an ending in *animate* nouns. But in the kind of system most commonly attested, ergative case is unmarked in *pronouns*, or in *first- and second-person pronouns*. Garrett takes this to be the result of subsequent analogical generalization.

Moreover, the lack of ergative marking in *inanimate* third-person pronouns is unexpected on his view and would require additional assumptions, perhaps analogy in the other direction within the pronominal system. And then, why do the putative analogical changes go precisely along the D-scale? Garrett appeals to morphological differences as a barrier to spread from nouns to pronouns, but that leaves unexplained the pronoun-internal hierarchy $1 > 2 > 3$.

And finally, if the "animacy" hierarchy in reality reflects a scale of topic-worthiness, (individuation, saliency, referentiality, etc.), as is widely recognized, then it is hard to see why this scale would have any *pragmatic* connection to the use of instrumental case.

This in no way detracts from Garrett's account of the origin of split ergativity in Anatolian, which is perfectly convincing, and still less from his methodological plea for a historical perspective in typological study. But this very historical perspective requires the appropriate theoretical underpinnings. Historical mechanisms by themselves cannot explain why languages undergo the particular kinds of reanalyses that result in split ergativity but not other, a priori equally imaginable kinds of reanalyses. The hierarchy in (15) must in some sense be part of the design of language. That still leaves the question whether this design is functionally grounded, structurally grounded, or both—an empirical question that is not decided by these data alone. A reviewer suggests that split ergativity may reflect universal preferences on what humans talk about and are interested in, which then lead to particular patternings of case marking as a result of paths of grammaticalization deriving from economy effects. On the other hand, the convergence of split ergativity with number and definiteness marking and especially the relation to agreement suggests a structural link with the determiner system.

Summarizing the discussion so far, the hierarchy (15) is a linguistic universal and SHOULD be expressed in the synchronic theory of grammar because:

(23) a. The hierarchy is inviolable.
b. There are multiple sources of split ergative case marking.
c. The hierarchy is a pathway of analogical change.
d. The hierarchy is manifested spontaneously in child language.

e. The hierarchy must be encoded in the grammar because it intersects with other hierarchies (notably definiteness) and because it plays a role in the distribution of other morphological categories (notably number and agreement).

2.3.3 Other manifestations of the D-hierarchy

If the hierarchy (15) is genuinely part of UG, it might be expected to play a role in other aspects of grammar than case marking. This proves to be the case. It interacts with other categories in an extremely revealing way.

The marking of NUMBER morphology reportedly follows the same hierarchy exactly:

(24) 'The singular–plural distinction in a given language must affect a top segment of the Animacy Hierarchy' (Corbett 2000: ch. 3, 4).

Thus, some languages distinguish singular from plural only in pronouns, and others distinguish singular from plural only in first- and second-person pronouns.

Also, any morphological marking of definiteness follows the hierarchy, for categories that distinguish definiteness (this excludes personal pronouns, which are inherently definite). Thus, a language may formally distinguish between definite and indefinite humans but not non-humans (for an example, see Old Georgian below), or between definite and indefinite animates humans but not inanimates.

Still more strikingly, agreement follows the hierarchy (15) exactly. The generalization is that it is more important to agree with first-person subjects than with second-person subjects and more important to agree with second-person subjects than with third-person subjects. This is shown by conjunction (*ego et tu sumus* 'I and you are.1PL' (Latin), etc.), and by various types of agreement systems (including, but not restricted to, inverse systems).

It has also been argued that preference for adjectival POSSESSORS follows the hierarchy (15) exactly (Koptjevskaja-Tamm 1993). An English example is the preference for *-s* genitives of pronouns and *of* genitives of nouns (Anttila and Fong 2003).

(25)	Its removal	?The removal of it
	?The tree's removal	The removal of the tree

The categories of number, definiteness, agreement, and possession are all related to the determiner system. Specifically, if the features of definiteness, number, and person are overtly marked in a language they are marked at least on its pronouns (and on its articles, if it has them). For example, any feature which participates in subject or object agreement in a given language is marked on the pronoun system in that language.

These generalizations, showing the relationship of the D-hierarchy in (15) to the determiner system, are interesting because they pose a problem not only for purely historical accounts of split ergative case marking, such as Garrett's explanation based on the derivation of ergative case from instrumental case, but also for functional accounts.

Recall that the functional explanation for why (15) governs split case marking is that high-D nominals occur more frequently as subjects and hence are left unmarked for economy reasons. It is mysterious why economy would demand the same hierarchy for number marking, or why the preference for specifier position would follow the same hierarchy. In fact, for agreement, the facts are the exact opposite of what the functional economy-based explanation would predict. It turns out to be *more* important, not *less* important, to mark subject agreement with the types of arguments which (according to the functional hypothesis) are most commonly used as subjects.

2.3.4 The basis of the D-hierarchy: a proposal

These data suggest that there may be a *structural* basis for the hierarchy that covers *all* its manifestations. It appears that ergative case marking may be incompatible with certain determiner features. In the simplest case, the structural link between the marking of case, number, possession, definiteness, and agreement is just the categorial distinction between nouns and pronouns/determiners.

The posited link between ergative morphology and the determiner system predicts that there should be languages where both ergative case and determiner features are overtly marked but may not co-occur. In what follows I present four examples which confirm this prediction.

In Old Georgian, ergative case marking was incompatible with morphological definiteness marking, including both number and specificity (Boeder 1979: 448).

(26)

	Non-definite	Definite	
		Sing.	Plural
Nominative	ø	-i	-ni
Dative	-s	-sa	-ta
Ergative	-man		

This correlation between case marking and definiteness suggests that ergative case was assigned to nominals headed by a noun, but not to nominals headed by a determiner. In other words, ergative case was assigned to bare NPs, but not to DPs.

In Koryak (Chukotko-Kamchatkan), ergative case marking is incompatible with morphological definiteness marking, which only occurs on human nouns. In (27), the distribution of the ergative marker *-a* reflects this grammatical restriction. Under these circumstances, ergative case is replaced by locative case, instrumental being

unavailable because it is restricted to inanimate nouns. Instead of ergative case, DPs are assigned instrumental case.

(27)

	Non-human	Human	
		Indef. stem	Def. stem
Instrumental	-a	—	—
Erg.	-a	-a	-k
Loc.	-k	-k	-k

If articles and pronouns belong to the same category D, and what is directly relevant to case assignment is category membership and feature content rather than semantics, then we would expect articles to group with the pronouns with respect to split ergativity. There are not many opportunities to put this prediction to a test, because ergative languages usually do not have definite articles.[8] One ergative language which does have articles is Arrernte (Australia). An article *re* (identical with the third-person pronoun) is postposed to NPs in a definitizing function. As expected, a transitive subject then receives an ergative case ending on the nominal but no case ending on the article (Andrews 2001: 10), as in (28).

(28) a. kngwelye re(*-rle) ker arlkwe-ke
dog the(*-ERG) meat eat-PAST
'The dog ate the meat.'
b. artwe re*(-nhe) kngwelye-le uthwe-ke
man the-ACC dog-ERG bite-PAST
'The dog bit the man.'

In Ngiyambaa, definiteness is marked on nominatives (absolutives) by a determiner cliticized to the preceding word. Definiteness is neutralized elsewhere, including in particular on NPs bearing ergative case (Donaldson 1980: 128), as in (29).

(29) a. mirigu=na bura:y gadhiyi
dog.ERG=3ABS child.ABSL bite.PAST
'The/a dog bit the child.'
b. mirigu bura:y gadhiyi
dog.ERG child.ABSL bite.PAST
'The/a dog bit a child/some children.'

Donaldson further points out that Kabardian (Northwest Caucasian) shows the same incompatibility between ergative case morphology and definiteness marking.

[8] Basque, which does, is not split ergative.

In Ngiyambaa, not only definiteness but also number marking is unavailable for ergative nominals. On the other hand, absolutive nominals which are marked as definite by a determiner are also obligatorily marked for number. Thus, *=na* in (29a) marks the object not only as definite, but also as singular; the corresponding plural determiner is *=naN-gal*.

A similar incompatibility between ergative case morphology and number marking is seen in Wargamay (Australian). According to Dixon (1981: 39–40), Wargamay regularly marks number only on pronouns which refer to humans. Corbett (2000: 55) summarizes:

> The first and second persons, singular, dual, and plural, and the third dual and plural are 'strictly specified for number' and are available only for reference to humans (and occasionally tame dogs). The form filling the third singular slot can range over all persons and all numbers (it can have non-human as well as human reference) but its basic sense is third person singular.

Nouns can be specified for number by reduplication, but this is restricted and not usual in the language; probably noun pluralization is not an inflectional process but a derivational one.

The Wargamay case-marking pattern is closely correlated with number marking. Nouns and the third singular pronoun (precisely the items which are inherently unspecified for number) are marked for ergative case only. The bare form without an overt case affix functions as an object and as a subject of intransitive verbs. Dual and plural pronouns in the first and second person are marked for accusative case only. The bare form without an absolutive is used in subject function. Finally, singular pronouns (which have no overt mark for number) have distinct ergative and accusative forms.

The generalization for Wargamay case assignment appears to be this:

(30) a. Elements that are inflected for number do not get ergative case endings.
b. Elements that cannot be inflected for number do not get accusative case endings.

2.3.5 Extending the account

Let us assume that the features of number and definiteness are registered in the determiner system, and that agreement is triggered by determiner features. Then the typological data reviewed above suggests that the assignment of ergative case is subject to the following constraint:

(31) Ergative case is assigned to projections of the category N, and not to projections of the category D.

This would account directly for what is by far the most common type of split ergative marking, which is between nouns and pronouns, on the usual assumption that pronouns and articles are members of the category D (Postal 1966, and much later work).

Split ergativity also commonly divides first- and second-person pronouns from third-person pronouns. My hypothesis then implies a corresponding categorical difference between them, at least in those languages. Such a difference has been proposed on independent grounds by Saxon and Rice, who claim that in Athapascan "first and second person, on the one hand, and third person, on the other, represent different morphological *categories*", registered in "different positions in the verbal complex", and (covertly) in "different hierarchical positions in the syntax" (Rice and Saxon 2005: 707).

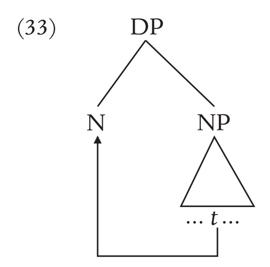
There are of course languages where pronouns do receive ergative case, and perhaps languages where DPs (nominals with definite articles) do receive ergative case. If we assume that (31) is universal, then in these languages it must be dominated by a more general constraint, perhaps one requiring every argument to bear case.

To explain the full hierarchy in (15), we can assume that DP can be headed not only by determiners and pronouns, but also, in some languages, by high-D nominals. (It was in anticipation of this assumption that I chose to refer to the hierarchy as the D-hierarchy). In some languages proper nouns and kin terms clearly function as Ds, either non-projecting or heading DPs. The evidence comes from inflectional morphology and from syntax. The morphological evidence comes from shared inflections between proper/kin nouns and pronouns (as in the above-mentioned Finnish child-language data). The syntactic evidence is a recurrent pattern of shared positional properties between proper/kin nouns and pronouns. Longobardi (1994, 2001) points out that kin terms and certain inalienably possessed nouns optionally have the distribution of determiners in Italian, as illustrated in (32).

(32)	Noi ricchi.	Mamma mia.	Gianni mio.	*Tavola mia.
	we rich	mother my	Gianni my	table my
	'We rich.'	'My mother.'	'My Gianni.'	'My table.'

Longobardi accounts for these data by head-to-head raising from NP to DP, and argues that the same pattern holds in a number of other languages.

(33)



Note however that the fundamental shared property of high-D elements on my proposal involves category membership; this may be reflected in syntactic positioning as Longobardi argues, but does not have to be.

Let us suppose that this process extends down the D-hierarchy in (15). This would not only account for the observed "animacy" effects on the marking of case, but would unify them with parallel asymmetries involving number and agreement.

At the same time, it gives us a new way to understand the apparent exceptions to the hierarchy in (15).[9] In Arrernte, the first-person singular pronoun, unlike all other pronouns of the language, receives the ergative case ending, like nouns (Andrews, 2001). This is surprising since first-person singular should be the very apex of the hierarchy, as demonstrated by the fact that in Dhalandji, ergative/nominative syncretism is reportedly restricted to just this pronoun (Austin 1981). It has been proposed that the Arrernte first-person singular pronoun is case-marked like a noun because a higher-ranked constraint requires it (Woolford to appear), but this stipulation does not relate its behavior to anything else. An alternative worth considering seriously is that it in fact *is* a noun. We know that the "personal pronouns" of some languages, such as modern Japanese, are really just nouns.[10] In particular, the Arrernte first-person singular pronoun might be a noun or DP, like Japanese *wata(ku)shi* 'I'.[11] There is no a priori reason to suppose that the translational equivalents of the English personal pronouns in any given language are morphologically and syntactically pronouns in it, any more than the translational equivalents of English auxiliaries or prepositions are necessarily auxiliaries and prepositions in it. This assumption gains some plausibility from the fact that the Arandic "pronouns" typically lack the morphology or phonology of function words, and they have nounlike syntactic properties. The duals and plurals have famously rich lexical meanings referring to patrimoiety and generation; they are also morphologically complex and may get their determiner properties from their suffixed number markers. (The third-person singular, however, is morphologically simple and clearly functions as a true pronoun/determiner; see *re* in (28).)

The upshot is that Arrernte's word for 'I' is a prima facie exception to (15), but if either Woolford's conjecture or mine should prove to be correct, it would not be a true exception; it would in fact support the interpretation suggested here, which relates the hierarchy to the syntactic and morphological category of its elements, over the traditional approach of tying it directly to meaning or reference. Syntactic and morphological evidence should resolve this question.

2.4 CODA NEUTRALIZATION

2.4.1 Final devoicing

Turning to phonology, let's begin with the robust phonological generalization that marked feature values tend to be suppressed in certain prosodic positions. Perhaps the best-known example of this process is CODA NEUTRALIZATION, the suppression

[9] For some examples, see Blake (2001) and Goddard (1982).

[10] "One acquainted only with modern Japanese would suppose that the language contained no true personal pronouns but only a number of periphrastic forms" (Sansom 1928: 71).

[11] Or for that matter like English "yours truly" or "the undersigned", which no one would call pronouns just because they refer to the speaker or writer.

of place and manner contrasts in syllable codas (or in word-final position), with the neutralized features typically taking their unmarked values. For concreteness let us consider the special case of devoicing of obstruents, as in German, most Slavic languages, Catalan, Turkish, Korean, and in a number of dialects of English.[12] As a sound change as well as synchronically, it seems to be irreversible, in that clear cases of the converse process of final voicing are not attested. Why should that be?

Again there are a priori two possible answers to this question. One locates the neutralization constraint in the design of language. This does not mean that coda neutralization applies in all languages; it just means that, whenever it does apply, it always imposes the unmarked feature value. It can be decomposed into two separate constraints. One says that onsets have at least as many place and manner contrasts as codas; which is really a special case of a family of constraints which differentiate between strong and weak positions. The other says that neutralized features assume their unmarked value (voicelessness, in the case at hand). This constraint too is more general. When voicing contrasts are neutralized elsewhere, the same generalization seems to apply. (Of course, we have to set aside contextual effects such as voicing assimilation, which override the default value and on some analyses take a neutralized archiphoneme as input.) The limiting case is context-free neutralization, i.e., the lack of a contrast in the phonemic system. Here the generalization seems to be that languages that have no manner contrasts realize their stops as voiceless and unaspirated (Maddieson 1984: 27).[13]

To say that the constraint is part of the design of language does not mean that it is arbitrary or unmotivated. For example, it has been suggested that the reason certain feature distinctions are liable to be suppressed in codas is that those feature distinctions are perceptually less salient in those positions (Steriade in press.)[14] As for why neutralization would favor unmarked feature values, the reason might be the greater economy of the relevant articulatory gestures. More effortful articulations would be used in positions where a contrast must be marked.

The second possible answer locates the explanation in the diachronic plane, along the lines of Blevins (2004a). Suppose that there are documented types of sound change that devoice final voiced obstruents, but there are no documented types of sound change that would voice them. It could then be claimed that the generalization follows trivially from this unidirectionality of change. Of course, the putative unidirectionality

[12] Australia, South Africa (Wissing and Zonneveld 1996). Many U.S. speakers, including the current president of the U.S.A., devoice word-final fricatives.

[13] This is *not* the same as saying that languages must have at least a voiceless unaspirated series of stops. In fact, they don't. Approximately 8 percent of languages in Maddieson's sample (1984: 27) are listed as lacking a series of plain voiceless stops; this includes English, where the voiceless series is enhanced by aspiration (as predicted by dispersion theory; see Flemming 2003; Steriade in press). Of course not all contexts in English have aspiration; it seems that Maddieson's analyses of phonological systems take something like the most common allophone of a phoneme as basic.

[14] Steriade however argues that syllable structure is not the correct characterization of the relevant positions. This is an important issue but I will have to set it aside here.

would in turn have to be explained; perhaps on the basis of the perceptual and articulatory asymmetries just mentioned, but now operating on the diachronic plane as constraints on the relevant sound-change processes.

The two answers are in many ways similar, and must ultimately converge in how they ground the asymmetry. The crucial difference between them is where the explanation applies. The historical account locates it solely in sound change. On this view, once a neutralization process has been "deposited" in a language by sound change, it just sits there as a brute arbitrary fact. It has no synchronic rationale, no more than, say, the geological stratification of the earth's crust does. The synchronic account, on the other hand, makes it part of the design of language. It says that a language which violates the universal (e.g., by having a final voicing process) would be not just historically unattainable, but synchronically complex, in the sense of being hard to learn, or hard to use, or both. Versions of OT phonology that posit that all constraints are universal and do not allow contradictory constraints even make the very strong claim that such systems are impossible.

Several considerations suggest that the historical account is not on the right track in this case. It is easy to construct scenarios that, unchecked, would produce the exact opposite process, of coda neutralization in favor of the marked feature value. For example, consider a language with a single series of obstruents with a gemination (or aspiration) contrast, neutralized in final position by degemination (or deaspiration, as the case may be). Let a lenition process then transpose the geminate/singleton (or aspiration) contrast into a voiceless/voiced contrast. Both these changes are common enough (the Romance languages offer many examples). With both sound changes taking place in sequence, the result would be final voicing. The data in (34) show the result of such a hypothetical change (with colons showing a contrast).

(34) A hypothetical path to final voicing: markedness reversal

	Medial			Final			
Stage 1:	atta	:	ata	at	:	(*att)	(final degemination, voicing not distinctive)
Stage 2:	ata	:	ada	ad	:	(*at)	(lenition)

The synchronic grammar at Stage 2 has a pattern which neutralizes voiced and voiceless obstruents as *voiced*.

Many other potential scenarios should be capable of producing the same result. Both voicing and devoicing are possible sound changes, and together they can reverse voicing.[15] In a language with an ordinary final devoicing process, such voicing reversals could theoretically produce synchronic final voicing processes, but that seems not to happen.

[15] Classical Armenian (and modern Eastern Armenian) T and D correspond respectively to modern Western Armenian D and T^h, (e.g., *Tigran* vs. *Dikran*). Germanic /θ/ and /d/ correspond respectively to German /d/ and /t/ (e.g., *Tod* 'death', *tot* 'dead').

In sum, final voicing *could* originate by everyday sound changes in a variety of plausible hypothetical scenarios. If it in fact never arises, *some constraint on the design of language must prevent it.*

2.4.2 Does Lezgian have final voicing?

Based on data in Haspelmath (1993a), Yu (2004) has suggested that there is at least one language that does have a syllable-final voicing process in its synchronic phonology, Lezgian (East Caucasian). The synchronic situation is that there are four distinct series of stops, of which three are invariant and the fourth alternates between a long voiced stop in coda position and a plain voiceless stop in onset position. This alternating stop series occurs only before the main stress, which falls on the second or only syllable of the word; it never occurs anywhere after the main stress.[16] The pattern is summarized in (35), where D = a voiced stop, D: = a long voiced stop, T' = an ejective, and T^h = an aspirate.[17]

(35)

	___V	V___]$_\sigma$
/D/	D	D
/T'/	T'	T'
/T^h/	T^h	T^h
/?/	T	D:

There is, then, a four-way manner contrast in pretonic position, and a three-way manner contrast elsewhere. The fourth stop series here labeled /?/ (the one which is restricted to pretonic position) is a plain voiceless stop in onsets and a long voiced stop in codas (where it contrasts with the non-alternating short voiced stop). The question is what /?/ is phonologically. Yu simply assumes without argument that /?/ is /T/, and that /T/ becomes voiced and lengthened in coda position. But there is really no reason why it could not equally be /D:/ which is conversely shortened and devoiced in onset position. The moraic theory of length tells us that /D:/ is really a geminate), so this solution amounts to positing onset degemination and onset fortition for Lezgian, which are both eminently natural processes. The three-series stop system /D/ : /T'/ : /T^h/ posited by this analysis is found elsewhere in the Caucasus (Georgian, according to some analyses at least, and Kabardian), and for that matter in Native American languages (Klamath, Kwakw'ala, Yana, Acoma) (Maddieson 1984).

This alternative, then, is typologically unexceptionable. Unless it can be excluded on compelling language-internal grounds, there is no case for coda voicing in Lezgian. And that means that the proposed universal is exceptionless as far as is known.

[16] "In a position following the stressed vowel, voiceless stops are always aspirated, never unaspirated, except [immediately after a voiceless obstruent, when they are always unaspirated]" (Haspelmath 1993a: 47). Haspelmath (1993a: 59) cites the form [t'^wek], which is inconsistent with the verbal rule just quoted; I assume it is a misprint for [t'^wek'] or perhaps for [$t'^w ek^h$].

[17] There is also an alternation between onset T' and coda D:, which reduces to the fourth set by a process which spreads the ejective feature, viz. ~ [T'VTV-] → [T'VT'V-] (alternating with [T'VD:]).

2.4.3 Further evidence for the universal

There is independent support for the conclusion that the markedness asymmetry seen in coda neutralization processes is not simply a by-product of sound change but reflects an intrinsic linguistic constraint.

First, the asymmetry arises in other ways than by the sound change of coda devoicing. In Konni (northern Ghana), regressive voicing assimilation (cf. (36a)) is blocked just when it would give rise to a voiced coda, in which case an epenthetic vowel is inserted instead (as in (36b)) (Cahill 1999).

(36) a. /tig-ka/ → *tikka* 'the village'
b. /biis-bu/ → *biisibu* 'the breast' (*[biizbu])

Similarly, in Meccan Arabic (McCarthy 2003), voicing assimilation (cf. 37a)) is blocked precisely when it would give rise to a voiced coda (as in (37b)), even though voiced codas are not in general prohibited in the language (see (37c)).

(37) a. /mazku:r/ → *masku:r* 'mentioned'
b. /ʔakbar/ → *ʔakbar (*agbar)* 'older'
c. /dabdaba/ → *dabdaba* 'sound of heavy footsteps'

Additional evidence comes from TETU ("The Emergence of the Unmarked") effects, manifestations of latent markedness constraints where higher-ranking constraints that override them are not in play (Prince and Smolensky 1993). For certain languages with strict CV syllable structure, final devoicing has been reported to apply spontaneously when speakers attempt to pronounce loanwords ending in -CVC. Language acquisition points in the same direction: final devoicing often occurs in the speech of young children (Smith 1973; Ingram 1976; Yavaş 1994; Wissing and Zonneveld 1996).[18]

No argument is needed to show that coda neutralization is a process which must be encoded in the grammar. It functions as a productive rule or constraint in numerous languages, interacting with other rules/constraints and principles within their phonological systems.

Thus, coda neutralization would seem to satisfy the requirements for a true linguistic universal.

2.5 STRESS/WEIGHT SOLIDARITY

Finally let's look at the well-known sonority hierarchy, and in particular the relative sonority of vowels, proposed by de Saussure and Jespersen (among many others) as an intrinsic universal, which is grounded articulatorily in the relative aperture of the vocal tract and/or acoustically in loudness and intrinsic duration.

[18] For example, Smith reports that his son Amahl, at age 2 years, 60 days, pronounced all stops (irrespective of their adult form) as voiceless unaspirated lenis initially, voiced in medial position, and voiceless finally.

(38) *a* > *e, o* > *i, u* > *ə*.

The relative sonority defined by (38) is one of a larger complex of sonority scales which involve syllable weight, pitch (de Lacy 2002b), and perhaps others. Several phonological constraints refer to sonority, but for our present purposes the relevant one is the stress/sonority solidarity generalization stated in (39).

(39) Stress seeks heavy syllables and sonorous vowels, where sonority is defined by the scale in (38).

De Lacy (2002a) extensively documents the generalization stated in (40), and argues that it is true of all universal hierarchies, at least in phonology.

(40) Adjacent points on the scale may be conflated, but not reversed.

Let us assume, for purposes of the following discussion, that de Lacy's empirical generalization is correct, and apply the criteria in (4) to determine whether the hierarchy (38) is truly a universal, or simply a typological generalization, as defined above.

As a basis for our discussion, let us consider the sonority-based stress system of Gujarati extensively analyzed by de Lacy (2002a), who formulates the empirical generalizations in (41).

(41)
- Words are normally stressed on the penult, but
- an antepenult is stressed if it is more prominent than the penult on the (partially conflated) sonority scale *a* > *e, o, i, u* > *ə*, and
- the final syllable is stressed if it is the only syllable with *a*.

The data in (42) show how stress is assigned in accord with (41).

(42) a. *azádi* 'freedom'
b. *ekóteɾ* '71'
c. *pəddhə́ti* 'guide'
d. *tájetər* 'recently' (*a* > *e*, attracts stress to the antepenult)
e. *kójəldi* 'little cuckoo' (*o* > *ə*, attracts stress to the antepenult)
f. *ʧɔpərá* 'girls' (*a* > other vowels, attracts stress to the final syllable)

De Lacy proposes for Gujarati a synchronic stress system which I informally summarize in (43):

(43) a. Assign stress to a heavy/sonorous syllable (STRESS-TO-WEIGHT), otherwise
b. assign a trochaic foot at the right edge of the word.

Couched in OT, de Lacy's proposal is explanatory in the sense that it predicts the possibility of the pattern in (41) from a set of ranked universal constraints.

The alternative historical explanation for the weight/stress solidarity seen in Gujarati might invoke the following reasonable assumption about sound change, in the spirit of Blevins (2004a):

(44) Intrinsic acoustic prominence of sonorous vowels may be reinterpreted as stress in sound change.

The validity of generalization (44) is not at issue here; let us assume that it is correct as stated. The question is whether it provides a sufficient alternative to de Lacy's proposal that (38), (39), and (40) are part of UG. Let us see what the criteria in (4) say.

The sonority hierarchy in (38) seems to be the same in all languages (this is the import of de Lacy's result in (40)). But there are natural types of sound change that *could* reverse it. For example, a sound change [ɑ] > [ə] (such as took place in Sanskrit, among other languages) could result in stress systems where ə functions as the *most* sonorous vowel, attracting stress in words like **tə́jetər*, contrary to (38)–(40). The result would be precisely a stress system with a rule/constraint which violates the proposed universal. The moral is that what sound change can create, it also can destroy. Therefore, if a generalization is exceptionless, there must be something more than sound change that sustains it.

Note carefully what is at stake here. The claim is *not* that sound change cannot subvert the phonological regularities of a language. For example, sound change could presumably conflate or delete vowels in such a way as to destroy the phonological regularities of Gujarati's stress system, with the result that it would have to be reanalyzed with lexically marked stress. What sound change apparently cannot do is to arbitrarily *reverse* the universal hierarchies. If so, then it follows sound change cannot be quite as "blind" as the neogrammarians thought. It must operate under the control of UG. Changes that subvert universals must either be blocked, or the system they appear to give rise to must be reanalyzed.

The second criterion in (4) concerns multiple paths. Weight/stress solidarity is in fact implemented in very diverse ways across languages. For example, in Finnish, it both makes unstressed diphthongs monomoraic in the lexical phonology, and prevents contraction of unstressed long vowels (Kiparsky 2003). Neither of these manifestations of it can be attributed to the perceptual confusion between weight and stress. This shows that the sound change theory of the origin of sonority hierarchy effects is not general enough.

Third, the generalization must be encoded in the grammar of Gujarati because it is productive. It locates stress in Gujarati loanwords, e.g., *sinemá* 'movie theater' (presumably from English *cinema*), which gets final stress in Gujarati by (41) because of the sonorous *a*. A similar point can be made for Finnish, where the sonority hierarchy determines fixed secondary stress in loanwords (Anttila 1997; Kiparsky 2003).

Fourth, the generalization underlies TETU effects. In Finnish, stem-final syllables receive secondary stress optionally in the lexical phonology, with a frequency that is proportional to the sonority of the vowel. The presence or absence of this stress triggers far-reaching allomorphy effects on the stem and the ending (Anttila 1997; Kiparsky 2003). At the phonetic level, lexically unstressed syllables, regardless of their sonority, receive a rhythmic stress which is phonetically indistinguishable from the

lexical stress—but which has no effects whatsoever on allomorphy. Therefore, the effect of relative sonority on lexical stress and on allomorphy cannot be attributed to misperception.

Fifth, the sonority hierarchy defines a pathway of analogical change. Anttila (1997: ch. 3) documents the analogical spread of long genitives (the forms in the right-hand column of (45)) in noun inflection over the attested history of Finnish. He shows that it follows the course in (45).

(45)						
	-i, -u stems	*líntuin*	>	*líntujen*	'birds'	16th century
	-e, -o stems	*péltoin*	>	*péltojen*	'fields'	19th century
	-a, -ä stems	*ákkain*	>	*ákkojen*	'old women'	20th century

Recall the weight/sonority solidarity generalization: sonorous vowels (such as *a*) prefer to be in stressed and heavy syllables, non-sonorous vowels (such as *i, u*) prefer to be in unstressed and light syllables. From this perspective, the trajectory in (45) is understandable: the morphological analogy changes heavy syllables into light syllables, and it is implemented first for high vowels, where weight/sonority solidarity favors it most, and last for (underlying) low vowels, where weight/sonority solidarity does not favor it at all. Here the sonority hierarchy governs the course of morphological change in a way which cannot have anything to do with the misperception of stress.

2.6 CONCLUSION

An increasingly popular research program seeks the causes of typological generalizations in recurrent historical processes, or even claims that all principled explanations for universals reside in diachrony. Structural and generative grammar has more commonly pursued the reverse direction of explanation, which grounds the way language changes in its structural properties. The two programs can coexist without contradiction or circularity as long as we can make a principled separation between true universals, which constrain both synchronic grammars and language change, and typological generalizations, which are simply the results of typical paths of change.

The following criteria should converge to identify true universals. (1) Universals have no exceptions (for what does not arise by change cannot be subverted by it either). That is, they are violable only in virtue of more highly ranked universal constraints. (2) Universals are process-independent. (3) Universals can be manifested in "emergence of the unmarked" effects. (4) Universals constitute pathways for analogical change. (5) Universals are embedded in grammars as constraints and can interact with other grammatical constraints.

Choosing as testing grounds Binding Theory and split ergativity in morphosyntax, and voicing neutralization and sonority in phonology, I argued that these criteria do

converge rather cleanly in each case. Pica's Generalization about the binding properties of simple vs. complex anaphors and Everaert's generalization that there are no nominative anaphors are not universals, but diachronically explicable typological generalizations. The D-hierarchy that governs split case assignment, number marking, and agreement is a universal, as is the direction of voicing neutralization and the sonority hierarchy in phonology.

3

On the Explanation of Typologically Unusual Structures

Alice C. Harris
State University of New York at Stony Brook

[Organisms are] bundles of historical accidents, not perfect and predictable machines.
Stephen Jay Gould (1983), quoted by Roger Lass (1990: 81)

3.1 INTRODUCTION

A classic problem in language typology is the explanation of why, on the one hand, some forms or constructions are common and others rare and, on the other, why the rare ones exist at all. As linguists, we need to explain (a) why unusual constructions are unusual (i.e. why they are found in few languages) and (b) if they are so unusual, why they can exist in some languages. Most explanations that have been proposed for the preference of one construction seem to be incompatible with the existence of the rare or unusual.

For example, on the one hand, why are SVO and SOV orders so frequent and OVS so rare, and, on the other hand, why does OVS exist at all? One "explanation" that has been proposed is that humans prefer to have S before O. But even if we accept that as

This paper was presented at the Explaining Linguistic Universals Workshop at Berkeley in March 2003, later that year at the Mediterranean Morphology Meeting in Catania, and at the Spring Workshop on Linguistic Reconstruction in 2004. Brief versions of the argument and of the Udi section of the paper were published as Harris 2005. The research reported here was supported in part by the National Science Foundation under grant BCS-0215523; gathering and analysis of data were supported in part by earlier grants, including a National Science Foundation National Needs Postdoctoral Fellowship (1978–9), the American Council of Learned Societies' exchange with the Academy of Science of the USSR (administered by the International Research and Exchanges Board, 1981, 1989), and National Science Foundation grants BNS-7923452, BNS-8217355, and SRB-9710085. I appreciate the support of these organizations. I am grateful to my Udi consultants, especially Luiza Nešumašvili, Dodo Misk'ališvili, Nana Agasišvili, and Caco Čik'vaiʒe. I also appreciate comments of all three audiences, especially those made by Adam Albright, Paul Kiparsky, Dan Slobin, and John Whitman.

an explanation, if it is true, why do speakers of Hixkaryana tolerate OVS as the basic order, with O before S? Do they have different preferences? If so, why?

In this paper I argue, with Lass in the article quoted above (1990), that languages too are "bundles of historical accidents". I argue further that unusual or rare features are unusual or rare because they are the accidental result of many different circumstances or conditions being lined up in just the right way.

In section 3.2 I discuss the structure of my argument and compare my explanation with other general approaches. Sections 3.3 and 3.4 describe a very rare structure in Georgian and in Udi, respectively, and show that each may be explained by the approach adopted here. In section 3.5 I provide a brief discussion of the Uniformitarian Principle in this context, and in section 3.6 I state conclusions.

3.2 THE STRUCTURE OF THE ARGUMENT

Before outlining the structure of the argument, I want to address what it means to explain. One kind of explanation consists of applying a well-understood principle to the unfamiliar (e.g. exotic constructions). If the exotic can be shown to be governed by well-understood principles, then it has been explained, in one sense of explanation. In this paper I show that some rare phenomena are rare because of simple probability of the historical changes and conditions required to establish them. I show that the development of some rare phenomena require many steps or conditions, and that it is rare for such steps and conditions to all coincide. The explanation is probability, a well understood principle.

Linguists could take the position that very rare phenomena are the result of the fact that

- our innate endowment discourages this (probably as part of some more general feature)[1]
- this system does not function well, perhaps because it is difficult to process
- this system is not acquired easily by children.

It may well be that each of these explains some phenomena,[2] but I see four general problems with each of these approaches to very rare structures. (i) In general, there is little direct evidence to indicate what our innate endowment provides; similarly there is little direct evidence that rare structures are dysfunctional. Difficulty of acquisition has often been equated with lateness of acquisition: if a structure is acquired late, that shows us that it is difficult to acquire. But when we consider complex structures

[1] Since the predominance of single systems of case marking is no more than a "typological generalization" in the sense of Kiparsky (this volume), some may not wish to consider the possibility that this is to be accounted for as part of our innate endowment.

[2] In particular, I have recently argued that innateness is the only way of explaining a set of similar changes that occurred in Tsova-Tush, Udi, and Georgian (Harris, in press).

such as relative clauses, which are not among the first things acquired by children, but which are by no means rare, it is not clear that it is always appropriate to link rarity with lateness of acquisition. (ii) All three views summarized in the bullet points above assume that independent evidence does exist. Because there is little or no independent evidence of this kind, the reasoning these views reflect is actually circular: one linguist may explain that certain structures are infrequent because they are innately discouraged/dysfunctional/difficult to acquire, while another tells us that the evidence that the structure is innately discouraged/dysfunctional/difficult to acquire is that it is infrequent. That is, without evidence independent of frequency, we cannot cite innate endowment, difficulty of acquisition, or functionality as an explanation without risk of circularity. (iii) If the innate endowment of human beings, or functionality, or acquisition makes certain structures less than optimal, we must explain why some are known to have endured for a long period of time. (iv) If any of these is correct, *we still have the task of explaining under precisely what circumstances our innate language capacity will permit a dispreferred system and under what circumstances it will not*.

I propose instead an explanation based on probability. If a construction can only develop by passing through a relatively large number of changes, or can only develop if certain conditions exist, or some combination of these, simple probability tells us that it will be less common than a construction that develops through fewer steps or requiring fewer conditions. This explanation does not depend on one change being less common than another, or on some conditions being infrequent; on the contrary, it assumes as a starting point that all changes and all conditions are equally common. It is the combination that is uncommon, not any of the specific elements.

Note that my explanation does not encounter the problems described above for the other proposed explanations. (i) There is independent evidence for the principles of probability. (ii) This explanation is not circular; proof of the reliability of probability does not rely on linguistic change. (iii) Because this explanation of rarity does not depend on a characteristic of language in general, it makes no prediction that a rare structure would have a short half-life. (iv) Unlike the other modes of explanation described above, probability of co-occurrence predicts both that some structures will be more common than others and that less common ones may exist.

Consider Joseph Greenberg's suggestion relating to probability, which is superficially similar to mine.

> In general one may expect that certain phenomena are widespread in language because the ways they can arise are frequent and their stability, once they occur, is high. A rare or non-existent phenomenon arises only by infrequently occurring changes and is unstable once it comes into existence. (Greenberg 1978a: 75)

Greenberg explains rarity of a phenomenon in terms of rarity of changes and instability, but this only sets explanation forward one step. Then we must ask why these changes are infrequent, and why this construction is unstable. My explanation does not rely on appeal to infrequency of changes or instability of constructions. Rather,

it assumes that the changes that produce the structure are common; it is only the combination that is uncommon, simply because it requires so many different steps or conditions. It is improbable that many languages would go through the nine or so processes and conditions that are described here.

An additional problem with Greenberg's approach is that it relies on an independent notion of stability, which we have no measure of. If "a rare . . . phenomenon . . . is unstable", we would expect the highly unusual phenomena described below to be unstable, yet each has been attested for more than a millennium and a half. Both are highly rare, yet both have been stable over the full period of attestation of the language at issue. Thus, Greenberg's assumptions are inaccurate, and his explanation does not advance our understanding. Although my explanation may appear to be similar, it is actually quite distinct.

My explanation makes the prediction that structures that develop in a single step, other things being equal, will be common among languages of the world, while those that require a large number of steps will be rare. This can be illustrated with the example of affixes. We may assume that suffixes and prefixes develop in a single step. (One might argue that each develops in two steps, first from an independent word into a clitic, becoming an affix as a second step. The difference is immaterial here.) Cross-linguistic evidence suggests that circumfixes usually develop from existing prefixes and suffixes (Harris and Xu 2006).[3] Because it assumes the prior existence of both a prefix and a suffix, a circumfix requires at least three steps, not necessarily in this order: development of a prefix, development of a suffix, linking the two as a single affix. This predicts, of course, that circumfixes will be less common than either prefixes or suffixes. (It makes no prediction, of course, about the interesting difference between prefixes and suffixes.)

3.3 GEORGIAN SPLIT CASE MARKING

3.3.1 Introduction

The unusual use of two or even three different case-marking systems in Georgian presents a challenge to linguistics. Splits of this kind are not, of course, unique to Georgian. For example, Hindi has a split according to tense-aspect (Butt 2001, among many other sources). Jacaltec is an example of a language with a split according to clause status (main clause vs. subordinate, Craig 1977). While split case marking is not unique to Georgian, it may legitimately be considered unusual; and the three-way split is probably legitimately viewed as highly unusual. As far as I am aware, the only other

[3] It is, however, known that in a language that already has circumfixes, it is possible for a new circumfix to develop similarly in a single step (Harris 2002b and sources cited there).

TABLE 3.1. Attested case patterns in Old and Modern Georgian

	Direct object	Subject of intransitive		Subject of transitive
		Inactive	Active	
Series I	Dative	Nominative	Nominative	Nominative
Series II	Nominative	Nominative	Narrative	Narrative
Series III	Nominative	Nominative	Dative	Dative

languages in which a three-way split has been described are sisters to Georgian. The split is exemplified in (1).

(1) Case marking in transitive verbs in Modern Georgian
Series I: k'ac-i k'lavs γor-s
man-NOM kills pig-DAT[4]
'The man kills a pig.'
Series II: k'ac-ma dak'la γor-i
man-NAR killed pig-NOM
'The man killed a pig.'
Series III: turme k'ac-s dauk'lavs γor-i
evidently man-DAT has.killed pig-NOM
'Evidently the man has killed a pig.'

A series, in the terminology of Kartvelian studies, denotes a set of verb tense-aspect-mood paradigms, the morphology that distinguishes them, and the special syntax that accompanies them, including the special case marking that is discussed here. (1) illustrates only the cases assigned in transitive clauses. Complete case patterns for Old and Modern Georgian are shown in Table 3.1. (The form of this table follows that used in Sapir 1917; today "inactive intransitives" are usually called unaccusatives, and "active intransitives" unergatives.)

We could suggest that other languages lack systems such as that in Table 3.1 because

- our innate endowment discourages this (probably as part of some more general feature)
- this system does not function well, perhaps because it is difficult to process
- this system is not acquired easily by children.

(i) I know of no direct evidence to indicate what our innate endowment tells us about case systems (even as an example of some more general principle). Similarly, I am not aware of direct evidence that the system is dysfunctional. The hypothesis that the system is difficult to acquire **has**, in fact, been tested. Regarding the difference between Series I and II, Imedadze and Tuite (1992), in their survey of work on child acquisition

[4] In verb glosses, the English translations in this paper use the feminine pronoun unless the example is taken from a text that requires the masculine, or unless a neuter is required by the context; affixes, pronouns, and clitics in Georgian and Udi do not actually distinguish gender.

of Georgian, conclude that "the presence of two distinct case-marking patterns used with different sets of verb forms does not present an especially difficult problem" (1992: 104). They continue, "The data . . . suggest that split case marking systems, however complicated they may appear to linguists and language students, may not be such insurmountable obstacles for children" (1992: 103–104).[5] (ii) All three views are based on the fact that most languages lack complex case systems of this sort, and the reasoning would necessarily go like this: systems like that in Table 3.1 are infrequent because they are innately discouraged/dysfunctional/difficult to acquire; the evidence for this is that they are infrequent. Clearly this reasoning would be circular. That is, without evidence independent of frequency, we cannot cite innate endowment, difficulty of acquisition, or functionality as an explanation without risk of circularity. (iii) If the innate endowment of human beings preferences a simpler case system, how do we explain the fact that most aspects of the system illustrated here have been attested for a millennium and a half? The long period over which the system has been attested also argues strongly against the functional explanation or the acquisition explanation. (iv) None of these views provides an explanation of why a few languages, such as Georgian, Hindi, Jacaltec, have more than one case pattern and other languages do not. If our innate endowment, for example, discourages such complexity yet permits this very complexity to surface occasionally (e.g. in Georgian, Hindi, Jacaltec), we still must explain why these particular languages are unaffected by our innate endowment, or by language function, or by the requirements of acquisition. We must still explain why acquisition or function or endowment explains some languages but not others.

Consider now a historical explanation. Using Georgian as my example, I argue that complex systems of this kind are rare because developing one requires so many steps, in the appropriate order. That is, the individual changes are all common changes of familiar types, and the circumstances are not rare; but the coincidence of their occurring together or sequentially, as required, happens infrequently. I argue that our innate endowment does permit systems of this complexity; we know this because they occur in human languages. They are complex, but not too complex for children to acquire them; we know this because children do acquire them and have done so for more than a millennium and a half. The system is not dysfunctional; we know this because it has continued to function for such a long time; it is the basis for a long and rich literary tradition, and it has stood up to political oppression and attempts to curtail its use. Thus, it is merely a historical coincidence that the circumstances that led to this complexity occurred in Georgian rather than in some other language, and it is probability that prevents these events all occurring together in a great many

[5] Series III encodes the evidential and is also used for the perfect. Regarding this, Imedadze and Tuite state that "the Georgian children in Kaxadze's study did not acquire the [first paradigm of Series III] until after age 3" (1992: 66) and they go on to describe another child who acquired it at age 2;3. To evaluate this more completely, it would be necessary to compare it with the acquisition of evidential-perfect in other languages.

TABLE 3.2. Reconstructed case pattern for the predecessor of Series II in Pre-Common Kartvelian

	Direct object	Subject of intransitive		Subject of transitive
		Inactive	Active	
Series II	Nominative	Nominative	**Nominative**	Narrative

languages. The system is the result of historical accident, just as biological organisms are, as observed by Gould in the passage quoted at the beginning of this paper.

In the remainder of this section I sketch the story of the development of complex case marking in Georgian, working in chronological order, and further address the issue of why this is a better explanation than others. Because the evidence for these developments is complex, there is not time or space to present it all; it can all be found in Harris (1985). I begin the diachronic story with the origin of the difference between Series I and II; this development is more complex than that involving the evidential Series III.

3.3.2 Development of a distinction between Series I and II

The earliest system that can be reconstructed on the basis of comparative and internal data is the direct predecessor of Series II; this is reconstructed on the basis of the verb morphology (Deeters 1930: 115; Čikobava 1942: 233, 1943, 1948; Pätsch 1952: 5; Kavtaraʒ e 1954: 13; Schmidt 1973: 115; Rogava 1975: 275; Boeder 1979: 460–463; Harris 1985: 95–103; Nebieriʒ e 1988). The morphology of verbs in Series I is similar to that in Series II, with the addition of extra markers in the forms of Series I (Harris 1985: 167–230). In addition, relics of Series II case marking can be found with two verbs in Series I in Old Georgian. That is, two verbs, both meaning 'know', occur in Series I (Harris 1985: 103–104) with narrative case subject and nominative case direct object, just as other transitive verbs do in Series II.

There is evidence that at the time when only Series II existed, Pre-Common Kartvelian case marking was ergative in the strict sense; that is, subjects of transitives were marked with the so-called narrative case (also known as the ergative case), while direct objects and subjects of all intransitives were marked with the so-called nominative case, as shown in Table 3.2.[6]

The evidence for this reconstruction is primarily syntactic relics found in several of the languages (Harris 1985: 108–133); other arguments are found in Harris (1985: 133–144). This could be considered Stage 1 in the development of the Kartvelian series.

Into the system reconstructed in Table 3.2 an antipassive construction developed, which required special marking in the verb form. Among the evidence that enables

[6] This view is also supported by statements in Deeters (1927: 21) and Šaniʒ e (1973: 483–484), though neither linguist puts it that way.

TABLE 3.3. Reconstructed syntax in Pre-Common Kartvelian

	Direct object	Subject of intransitive		Subject of transitive
		Inactive	Active	
Productive antipassive	**Dative**	**Nominative**	**Nominative**	**Nominative**
Series II	Nominative	Nominative	Nominative	Narrative

us to reconstruct the antipassive as a productive construction is ablaut, which distinguished transitive from intransitive in the verb (Harris 1985: 167–188, drawing on the brilliant work on ablaut by Gamq'reliʒ e and Mač'avariani 1965), the use of markers of the collective to signal the imperfective antipassive (Harris 1985: 189–208), the distribution of the plural object marker, *-en*, in Old Georgian (Harris 1985: 209–230), the case marking found in Series I as an exemplar of case marking in antipassives generally (Harris 1985: 231–256), and the distribution of person agreement in Old Georgian (Harris 1985: 257–263). The resulting system is summarized in Table 3.3 and may be considered Stage 2 in the process described here.

Following this, the case marking in Series II changed from a true ergative system to the system sometimes labeled active-inactive, in which the subjects of active intransitives (unergatives) are marked like subjects of transitives, with the narrative case, while subjects of inactive intransitives (unaccusatives) are marked like direct objects, with the nominative case. Traces of the older system can be found in Old Georgian, and in this way we can track the transition (Harris 1985: 329–361). This system is summarized in Table 3.4 and may be considered the third stage in the changes described here.

Although the antipassive originated as a synchronically derived construction, the change reflected in Table 3.4 forced it to be reanalyzed as a non-derived imperfective.[7] After reanalysis, this construction retained the special case marking and much of the special verb morphology and is known now as Series I. This is summarized in Table 3.5 and may be considered Stage 4 in the development of complex case marking in Kartvelian.

Under somewhat similar circumstances, the construction that originated as an antipassive displaced the older construction in Lardil, a language of the Tangkic subgroup of the non-Pama-Nyungan languages of Australia (Klokeid 1978; McConvell 1981; Evans 1995), and in the Ngayarda languages of Western Australia (Dench 1982). As a result of a complex set of changes in Lardil, the parallel of Series I in effect replaced that of Series II and continued as the only system of case marking in the language. Why did this not happen in Kartvelian? In Lardil, the syntactic changes may have been triggered by a series of phonological processes that destroyed tense

[7] Although I have presented Stage 3 as preceding Stage 4, in my view this ordering is only a logical one, and it is most likely that the two changes occurred simultaneously, still with the reanalysis of Series I as a consequence of the change in case marking in Series II.

TABLE 3.4. Syntax reconstructed for Pre-Common Kartvelian

	Direct object	Subject of intransitive		Subject of transitive
		Inactive	Active	
Productive antipassive	Dative	Nominative	Nominative	Nominative
Series II	Nominative	Nominative	**Narrative**	Narrative

marking (McConvell 1981). Both McConvell (1981) and Evans (1995: 423–450) argue than an antipassive and other accusative patterns replaced the old ergative main clause pattern, though they give somewhat different accounts of the process. In fact, phonological erosion of the ergative marker occurred in Kartvelian, though it probably happened relatively later in its history and had no effect on the changes at issue here. When this did occur, Kartvelian had a ready source of new case markers, which it utilized in renewing its narrative (ergative) case (Harris 1985: ch. 4). Dench (1982: 54) observes that the motivation for the change in Proto-Ngayarda is not clear, but he suggests that it may have served the purpose of identifying the syntactic pivot. It is possible that full replacement of the ergative-absolutive case system with a nominative-accusative one did not take place in Georgian because neither of these motivations, nor any other, was present. In addition, in Kartvelian there was so much redundancy in the morphology that related yet distinguished the two series of verb forms and the case patterns that co-occurred with them, that both could survive. Similar survival and co-occurrence of an input form and a reanalyzed form are known from attested changes in Kartvelian and many other languages (Harris and Campbell 1995, 1996).

It is most likely that the crucial difference between the Australian languages cited above and Common Kartvelian in this regard is the Kartvelian reanalysis in Series II, which resulted in the system sketched in Table 3.4, and the related change in the use of cases in derived intransitives. At the time of the innovation of the antipassive, subjects of derived intransitives of all types were marked with the nominative case (see Table 3.3). But by the time of Old Georgian, making a direct object an oblique, as in antipassives, did not result in the subject being marked with the nominative case (Harris 1985: 330–342). In fact, one may assume that a language with the case-marking

TABLE 3.5. Syntax reconstructed for Common Kartvelian (and attested in Georgian and Svan)

	Direct object	Subject of intransitive		Subject of transitive
		Inactive	Active	
Series I	Dative	Nominative	Nominative	Nominative
Series II	Nominative	Nominative	Narrative	Narrative

system summarized for Series II in Table 3.4 could not logically have an antipassive of the kind summarized in Table 3.3, and I know of no evidence that such a language exists.[8] Thus, the change in Series II explains the reanalysis of Series I (see note 7). After the case system changed from a true ergative one to one in which subjects of intransitives were marked in two different ways, the case marking of the antipassive could no longer be predicted from the fact that the construction was intransitive.[9] The fact that it was no longer possible to relate the syntax of the two series led speakers to reanalyze them as independent (sub)systems.

3.3.3 Development of the distinction between Series I/II and Series III

The contrast between Series I and II, on the one hand, and Series III (the evidential), on the other, is more recent and much simpler. Its origin is complex, in the sense that the morphology of some paradigms comes from Series I, while that of other paradigms comes from Series II (Harris 1985: 286–292, and sources cited there), and in the sense that some new morphology had to be developed (Harris 1985: 293–295, and many Georgian sources cited there). As an example of reuse of Series I morphology, compare the Series I (lexical) inversion in (2a) with the morphology of the Series III inversion in (2b). ("Inversion" is a term traditional in the study of languages of the Caucasus. The inversion construction is one in which experiencers and in some instances other nominals that are expected to be subjects have syntactic and morphological properties of indirect objects, while stimuli (themes), if present, have properties of subjects.)

[8] At the workshop at Berkeley, it was suggested to me that Bandjalang, as described by Crowley (1978), is indeed a language with active-inactive case marking of the type summarized for Series II in Table 3.4 and an antipassive of the type summarized in Table 3.3. This is, however, incorrect. According to Crowley, intransitive subjects generally are marked -Ø (1978: 52); this is not like an active-inactive case system, where intransitive subjects are marked differentially—some in one case, and others in another. Austin (1982), drawing on Crowley's data, calls attention to a group of eight verbs which behave syntactically like transitive verbs, taking an ergative subject, but which never take an overt direct object (see also Crowley 1978: 107–108). Both authors also cite one example (the same example) showing that this verb can occur in the antipassive construction. Harris (1981: 181–190) discusses in detail a small group of similarly exceptional verbs in Georgian; these verbs, too, behave syntactically as though they had a covert object, and examples are given in a range of constructions to show this. This group of verbs was somewhat larger in Old Georgian, and Harris (1985: 329–361) documents the role of these verbs in the transition to the system shown in Table 3.5. However, in contrast, there are literally hundreds of regular verbs that take ergative subjects in Georgian but are syntactically intransitive as shown in Harris (1981, *passim*); most of these are listed in Holisky (1981). Harris (1985: 331) notes that in languages with true active-inactive case marking, there is no change of the case marking of the subject in derived intransitives, citing Onondaga as an example.

[9] Harris (1981: 146–150) shows for Modern Georgian that in order to derive Series I from Series II by antipassive or any other rule or relation, one would have to state each of the following rules twice, once for Series I and once for Series II: subject person agreement, direct object person agreement, object camouflage, unemphatic pronoun drop, and number agreement. Most of the same arguments can be made for Old Georgian and the sister languages.

(2) Old Georgian

a. Series I inversion		b. Series III inversion		c. Series I (non-inversion)	
m-i-pq'r-ie-s	'I hold it'	m-i-c'er-ie-s	'I have written it'	v-c'er	'I write it'
g-i-pq'r-ie-s	'you hold it'	g-i-c'er-ie-s	'you have written it'	s-c'er	'you write it'
u-pq'r-ie-s	'she holds it'	u-c'er-ie-s	'she has written it'	c'er-s	'she writes it'

The morphology that sets apart the older lexical inversion, illustrated in (2a), includes the use of a prefix *i*-, together with a person prefix *m*- or *g*-, or the use of a portmanteau morph, *u*- in the third person. These are markers otherwise characteristic of certain indirect objects in Georgian (Harris 1981: ch. 6). The older inversion forms are also characterized by the suffix *-ie* (apparently derivational in function), and the suffix *-s* marking agreement with the theme nominal; the latter otherwise indicates agreement with a third-person subject in certain tense-aspect-mood categories. The syntax that sets apart the older lexical inversion is the treatment of the experiencer as indirect object and treatment of the theme as subject. The innovation described here was putting this morphology and syntax to use with transitive and unergative verbs that did not (and do not) govern lexical inversion. This resulted in the complex system found today, repeated in Table 3.6.

Aramaic is a language that has undergone a parallel change; older stages of Aramaic had a construction (with the so-called J stem) in which the experiencer-nominal was expressed with oblique pronouns, which originally had been limited to indirect object, then were extended to direct objects. The stimulus-nominal (or theme) in the same construction was expressed with the so-called A-set suffixes, which originally had been the nominative set of pronouns (Hoberman 1988). An important difference between Georgian and the dialect of Modern Aramaic described by Hoberman is that in the latter both sets of pronouns have become affixes, whereas Georgian has both affixal agreement and independent nouns or pronouns.

3.3.4 Discussion

Thus, development of the contrast between Series I and II in Georgian involved the following factors, possibly together with others. (i) **Ergative case marking** in early Pre-Common Kartvelian. The antipassive construction generally occurs only in languages with true ergative case marking or agreement or both (cf. Dixon 1994, esp. pp. 17

TABLE 3.6. Attested case patterns in Old and Modern Georgian

	Direct object	Subject of intransitive		Subject of transitive
		Inactive	Active	
Series I	Dative	Nominative	Nominative	Nominative
Series II	Nominative	Nominative	Narrative	Narrative
Series III	**Nominative**	**Nominative**	**Dative**	**Dative**

and 146–152). (ii) **The development of the antipassive**. Although it appears that the antipassive can exist only where there is ergative morphology, these are independent features of a language. That is, it is not the case that every language with ergative morphology has an antipassive, so the development of this construction is a distinct characteristic of early Common Kartvelian. (iii) **Change in the Series II case-marking system from true ergative to active-inactive**. This change is probably crucial to causing the split, for otherwise the antipassive could continue to be derived productively from the active form. (iv) **Reanalysis of the antipassive**. It is the reanalysis that provides the actual split. (v) **Abundant morphology**. It is likely that maintaining the contrast between Series I and II has at various stages also depended on (a) ablaut, (b) series markers, and (c) object agreement, each of which either registers the differences between Series I and II or tracks the grammatical roles of arguments. Together this morphology made it possible to distinguish the antipassive (later Series I) from the Series II forms out of which they developed. Perhaps other morphology could have done the job as well, but it is likely that there had to be some morphology to perform this function. The development of Series III involved the following: (vi) **Lexical inversion**. Clearly there are languages, such as French, that lack lexical inversion. Yet it is found in many languages of India (see Verma and Mohanan 1990), most languages of the Caucasus, and in other languages of the world. (vii) **Verbal morphology indicating inversion**. Inasmuch as some languages, such as German, have lexical inversion without special verb morphology, this is an independent characteristic.[10] (viii) **Overt case marking**. Given that neighboring Abxaz has lexical inversion with no overt case marking, this too is an independent variable.[11] (ix) **Development of an evidential**, probably as a result of areal pressure (Friedman 1979). Thus, there were at least five circumstances or processes that coincided to permit the development of the Series I/II contrast and an additional four to develop the Series III contrast. None of these nine factors is especially unusual; indeed, each is widespread, either in the region or in the world at large. It is the coinciding of all nine factors that is uncommon. Thus, we can say that neither German, nor O'odham, nor Chichewa, nor Thai has a case split of the

[10] For example, the German expressions below involve inversion and require no special verb morphology.

(i) mir ist kalt
me.DAT is cold
'It is cold to me' or 'I am cold.'

(ii) mir scheint das gut
me.DAT seems that.NOM good
'That seems good to me.'

[11] Ketevan Lomtatiʒ e, lectures, University of Tbilisi, 1974. An Abxaz example transliterated from Lomtatiʒ e's lectures follows.

(i) y-s- mo-wp'
3SG.N.ABS-1SG.IO-have-COPULA
'I have it.'

It is in part the position of the two markers that distinguishes their functions, since *s*ə- also marks first persons, for example, in other functions in other positions.

type that is found in Georgian because neither these nine factors nor others that might have led to the same result happened to coincide in those languages.

But couldn't the same system have developed through a different chain of events, possibly simpler? The only similar systems we know of developed in a roughly similar way. While there are parallels in some Indo-Iranian and some Mayan languages for the development of systems similar in some ways to Georgian's Series I and II, to the extent that the history of their development is known, it is not simpler (see, for example, Anderson 1977 and Payne 1979 on Indo-Iranian languages). In other instances active-inactive marking may develop in other ways, but it is unlikely to be simpler. This occurred in Tsova-Tush, for example, and involved development of the differential case marking on pronouns, cliticization, then affixation, and thus must have been more complex than the relatively simple process in Georgian. As mentioned above, Aramaic provides an example of the reanalysis of the construction called inversion in the Caucasus, but this involved the development of new morphology and was not significantly simpler than the related changes in Georgian. While we have parallels for each of these three changes, and of course parallels for other circumstances such as abundant morphology, lexical inversion, etc., other languages do not point the way to simpler ways of developing any one of the three subsystems found in Georgian. Further, it is difficult to imagine simpler routes that would lead to these three subsystems.

Some might complain that the historical account fails to explain why Georgian has not "repaired" the system, as the sister language Laz has to a great extent. Laz had a case system like that in Georgian but made it more regular through two changes: (i) It simply began to use the Series II case-marking system with Series I verb forms. (ii) A comparably simple change in Series III would probably not have been possible because in Series III verb forms, the dative-nominal conditions indirect object agreement, not subject agreement, even though it is the agent. Instead, Laz developed new verb forms with the evidential meaning of Series III, and the old forms, together with their syntax, have nearly disappeared. The idea of repair here is not necessarily ethnocentric, but is based on regularity. "Why can't Georgian be more regular? Why can't it make do with one case system?" This is like asking why Russian has to have so many different forms for the dative case (Georgian has only one), or why English has more than eight forms for expressing the future (why can't it make do with one?).[12] The answer for Georgian is that there is apparently no need of repair: the system works and can be acquired. Although many languages simplify, there is nothing about our innate endowment that demands that a language simplify.

[12] I have in mind here the simple present (*Tomorrow I wash the car*), present progressive (*Tomorrow I am washing the car*), simple future (*I will wash the car*), *Be going to* (*I am going to wash the car*), *Be to* (*I am to leave at 8 o'clock*), future progressive (*I will be washing the car*), future perfect (*I will have washed the car*), future perfect progressive (*I will have been washing the car*).

We have seen, then, how historical accident can explain the fact that an unusually complex case system developed. Note, however, that the historical analysis not only accounts for the complexity *per se* but also accounts for the particular case-marking patterns found. By relating the Series I syntax of Kartvelian languages to the antipassive of other languages with some ergative morphology, this analysis explains the specific case marking found. In the antipassive in other languages, direct objects are demoted or otherwise cease to be direct objects, leaving all verbs intransitive. Correspondingly subjects are marked with the intransitive subject case, the nominative (absolutive) case. The demoted direct object itself is marked as an indirect object in the antipassive. In languages of other families, the marking of the initial direct object may vary somewhat, but it is generally marked with some case other than that used for subjects and direct objects. For example, in Yidiɲ, in an antipassive the initial direct object is in the dative (Dixon 1977: 274).[13] Among languages of the world that have an antipassive, the construction indicates imperfective or durative aspect and contrasts with a perfective or punctual, just as it did in Common Kartvelian.[14]

Series III, too, has the case marking shown in Tables 3.1 and 3.6 and illustrated in (1) because of its origins. In every Kartvelian language, and indeed in many other languages of the world (see Verma and Mohanan 1990), a set of affective (or psych) verbs lexically governs inversion in all series; that is, they take dative experiencers and nominative stimuli obligatorily or optionally. Some cognates form the basis for reconstruction of this syntax in Common Kartvelian, as illustrated for Series I in (3).

(3) a. Svan
ali segz-s ka xesmi (√sm)
this.NOM Segz-DAT pv he.hear.it
'Segz hears this.' (Davitiani et al. 1957: 10, 37)

b. Mingrelian
ragadi masime(n) (√sim)
talk.NOM I.hear.it
'I hear the talking.' (Čikobava 1938: 314)

[13] For example, consider the sentences below from Dixon (1977: 274).

(i) waguḍa-ŋ gu buɲa giba:l — not antipassive
man-ERG woman.ABS scratched
'The man scratched the woman.'

(ii) wagu:ḍa giba:ḍiɲu buɲa:-nda — antipassive
man.ABS scratched woman-DAT
'The man scratched [at] the woman.'

[14] Other linguists have made the same observation, notably Zorrell (1930: 93), who compared Georgian Series I with German *er malt an einem Bild* (in contrast to *er malt ein Bild*), and Anderson (1977: 349–352), who compared Series I with *John shot at Bill* (in contrast to *John shot Bill*). See Harris (1985: 159–161) for a complete description of the contributions of other linguists. On the use of the antipassive "to refer to activity that has not actually been carried out or has not been carried through to completion", see Blake (1977: 16) and Dench (1982: 54).

c. Modern Georgian
vis esmis čemi lap'arak'i? (√sm || sem)
who.DAT he.hear.it my talk
'Who hears my talking?'

On the basis of examples of this type, we can reconstruct the inversion construction in Series I and II to Common Kartvelian with some specific affective verbs.[15] The inversion construction with affective verbs marks an action as unintentional or involuntary; this can best be seen with those stems that are optionally affective. For example, direct *vič'er* is 'I catch it', while inverted *mič'iravs* is 'I have, hold it'. Unintentionality is one component of the meaning of Series III in all Kartvelian languages (see Harris 1985: 280, 281, 288).

(4) Modern Georgian
a. merab-s ar unaxia Series III
Merab-DAT NEG he.see.it
'Merab has not seen it.'
b. merab-ma ar naxa Series II
Merab-NAR NEG he.see.it
'Merab didn't see it.'

(4b) implies that Merab refused to see it, while (4a) is neutral in this respect. Other components of the meaning of Series III—perfect aspect and evidential mood (that the speaker did not directly witness the action or state)—also express a distancing of the subject or speaker from the action.

This provides an explanation of the complex case system of Georgian in the sense that it attributes the infrequent occurrence to the improbability that some nine changes or conditions (or others equivalent to them) would occur often in languages of the world. This explanation does what no other proposed explanation can do—it shows both why a complex case system of this sort occurs infrequently and why it occurs at all.

3.4 UDI ENDOCLITICS

3.4.1 Origins

Udi is unrelated to Georgian, belonging to the Northeast Caucasian (NEC) family of languages.[16] In Udi the typological problem to be explained is the unusual existence of endoclitics. They may occur intermorphemically, as in (5), intramorphemically (inside

[15] Šerozia (1980) has argued that the syntax of this construction is to be reconstructed to CK also for a potential construction, not found in attested stages of Georgian.

[16] Udi is a member of the Lezgian subgroup of NEC, which also contains Lezgi, Tabassaran, Aγul, Rutul, C'axur, Budux, Kryz, Arči, and possibly Xinalug.

a single morpheme), as in (6), enclitic to the verb (in the second clause in (6)), or enclitic to the focused constituent (7).

(5) qa usen-axo-ośa jesir pasčaγ-en xoyš-**ne**-b-sa me pasčaγ-ax te . . .
20 year-ABL-after captive king-ERG request-**3SG**-DO-PRES this king-DAT that . . .
'After twenty years the captive king begs this (i.e. the second) king . . . ' (D 67: 2)[17]

(6) jesir pasčaγ-a bu-**t'u**-q'-sa ič ölkin-ä ta-γ-a-**ne** (D 67: 7)
captive king-DAT want$_1$-**INV3SG**-want$_2$-PRES self land-DAT thither-GO-SBJVI-**3SG**
'The captive king wants to go to his (home)land.'

(7) un ek'a-**n** maslahat-b-esa?(D 67: 8)
you.SG what-**2SG** advise-DO-PRES
'WHAT do you advise?'

In the verb in (5), the (general-purpose) third-person singular subject clitic, =*ne*, occurs between the light verb, -*b* 'do, make', and the element it incorporates, *xoyš* 'request'. In (6) we see the form *bu-t'u-q'-sa* 'he wants it' (a lexical inversion verb), where the third-person singular inversion subject marker, = *t'u*, occurs inside the synchronic root *buq'*- 'want'. In (7) the second-person singular subject clitic occurs outside the verb and its presence marks its host, *ek'a* 'what', as being in focus (indicated in the translation by upper-case letters). These clitics, =*ne*, =*t'u*, and =*n*, mark person–number subject agreement and are referred to in what follows as Person Markers (PMs). The morphemes at issue are not infixes, since they meet Zwicky and Pullum's (1983) criteria for distinguishing clitics from affixes and other definitions of "clitic" in the literature (see Harris 2002a: 94–114). The most important among these, in my view, is that clitics may occur on a variety of hosts; and the Udi clitics at issue occur on verbs, nouns, adjectives, adverbs, pronouns, or postpositions. According to a variety of criteria established in the field, the verbs within which these endoclitics occur are words, not phrases (Harris 2002a: 76–87). Similarly, the morpheme *buq'*- 'want' is a single morpheme by all synchronic criteria (Harris 2002a: 72–75). This is not what we expect from a typological point of view; how can we explain its occurrence here?

As possible explanations we might suggest that other languages lack morpheme-internal clitics because

- such clitics make the host morpheme difficult to understand (to process)
- our innate language capacity makes intramorphemic clitics difficult/"expensive"
- this system does not function well
- this system is not acquired easily by children.

However, (i) no **direct** evidence exists to support any of these alternative views. (ii) All four views are based on the fact that most languages lack word-internal clitics, and the reasoning would thus be circular. (iii) This system has endured for at least

[17] Examples designated "D" are from Dirr (1928). Examples with no designation are from my fieldwork.

1,600 years.[18] (iv) None of these views provides an explanation of why Udi (and presumably a few other languages) deviates from the common pattern. That is, if we accept any one of these views, we still must explain how Udi manages to maintain this pattern. For example, if we suggest that this construction is difficult to acquire, what makes it difficult, and why does this matter less for Udi than for other languages? Or, if we believe that endoclitics make the host morpheme difficult to process, do we explain their occurrence in Udi by suggesting that Udis are smarter or otherwise better at language processing than the rest of us? In contrast, the historical account of the infrequency of the constructions in (5–7) (i) is supported by the facts, although the evidence for some parts of this explanation is scant, (ii) is not circular, (iii) explains both why endoclitics are infrequent and why a few languages deviate from the common pattern.

Udi possesses an unusual construction because by accident its history holds an unusual combination of circumstances and events. Again, I cannot recount all of the details here, but the full account can be found in Harris (2002a). Here the story begins with a focus cleft. A number of NEC languages have a focus cleft similar to that found in the literary dialect of Dargi illustrated below (examples from Kazenin 1994, 1995; see also Kazenin 2002).

(8) a. x'o-ni uzbi arkul-ri
2SG-ERG brothers.ABS bring.PAST-2SG
'You brought the brothers.'
b. x'o saj-ri uzbi arku-si
2SG.ABS FM[COP-2SG] brothers.ABS bring-PTCPL.SG
'YOU brought the brothers.' 'It was YOU that brought the brothers.'

(9) [$_S$ FocC$_i$ Copula-Agmt$_i$ [$_S$... Verb]]
SUBJ PTCPL

We do not know whether Proto-NEC had focus clefts, but the construction is common in the family, whether through inheritance or through borrowing or through other means. There is reason to believe that Pre-Udi, like many of its sisters, had this construction. It is likely that in Udi the embedded clause was introduced with a pronoun, which is represented as 'that$_i$' in (10).[19]

(10) [FocC$_i$ Copula [$_S$ that$_i$. . . . Verb]]
SUBJ PTCPL

[18] Udi was written as early as the fourth or fifth century AD, and it has been known for some time from a very small number of inscriptions. However, in 1975 palimpsest texts of older Udi (also refered to as Caucasian Albanian) were found in the St. Katherine Monastery on Mt. Sinai, and work was begun on them by Zaza Aleksidze. Some twenty years later, Aleksidze began to work with Wolfgang Schulze, Jost Gippert, and Jean-Pierre Mahé, and publication is promised soon (Aleksidze et al., in preparation). Schulze (2004, 2005) reveals a few facts about this older form of Udi, including that it already contains intramorphemic clitics. The Old Udi texts have not yet been definitely dated, but Schulze (2005) estimates their dates as fifth to eighth centuries.

[19] In the third person, the pronoun was probably *t'e* 'that, yon' (see Harris 2002a: 239).

Clefts are characterized by an embedded clause expressing a presupposition and containing a gap or variable. In Udi, the pronoun 'that$_i$' represented the variable in the embedded clause and was coreferential with the focused constituent in the matrix clause. We may consider (10) to represent Stage 1 in the development of the patterns illustrated in (5–7). The copula in (10) may have been null at the time the cleft construction was developed or inherited; if not, it was dropped at some point.

(11) [FocC$_i$ Ø [$_S$ that$_i$. . . . Verb]]
SUBJ COPULA PTCPL

The biclausal structure was reanalyzed as monoclausal in accordance with the universals of syntactic changes of that type (Harris and Campbell 1995: ch. 7); the modern PM is the reflex of *that*$_i$ in (11).

(12) [FocC -PM . . . Verb . . .]
FINITE

The transition from (11) to (12) included the reanalysis of the focused constituent as subject of the lexical verb; this in turn occasioned a change in the case of this subject, and a change in the form of the verb from participial to finite. This explains the origin of the focus construction illustrated in (7). There is only rather circumstantial evidence to support this much of the story, and hence we can be less certain of it than of what follows.

In some instances, sequences of focused constituents (FocC) and the verb that followed them, as in (12), were reanalyzed as verbs with an incorporated noun (or other element), a process referred to here as univerbation. Since the FocC was always followed by the PM, the latter was trapped in the process of univerbation, as shown in (13).

(13) FocC-PM V-TAM > IncE-PM-LV-TAM Pre-Udi

The change represented in (13) accounts for the verb form in (5) above, as shown in (14).

(14) IncE - PM - LV - TAM from (5)
xoyš- ne- b- sa

For some lexical items, a contrast is maintained between items that have undergone this univerbation and otherwise identical ones that have not, as in (15).

(15) a. FocC-PM V-TAM q'ulluγ-ne b-esa Udi
service-3SG DO-PRES
'she does SERVICE'
b. IncE-PM-V-TAM q'ulluγ-ne-b-(e)sa
'she serves'

Over time, most of the monomorphemic verbs in Udi have been lost, leaving mostly complex verbs, such as those in (14) and (15b). On this basis, it became expected that, in

the absence of focus or other special circumstances, PMs would occur as endoclitics.[20] That is, in typical complex verbs, such as those cited here, the PM immediately precedes the light verb, the final consonant of the stem. The stems of monomorphemic verbs are generally of the form (C)VC(C). By moving to the position before the final consonant, the PM began to occur in the "same" position in monomorphemic verbs as in complex ones. This change is represented in (16), with the verb root *uγ* 'drink' (see Harris 2002a: ch. 9 for details).

(16) IncE - PM - LV - TAM || ROOT$_1$ - PM - ROOT$_2$- TAM
xoyš- ne- b- sa *u- ne- γ-* sa 'she drinks'

The symbol "||" may be read here as "is the basis for the analogical creation of the new form". The element labeled ROOT$_2$ is always a single consonant, like the majority of light verbs; the element labeled ROOT$_1$ may consist of a single vowel, as in this example, or of CV or CVC. The schema in (16) accounts for the creation of the form *bu-t'u-q'-sa* 'he wants' in (6), from the root *buq'-* 'want'.[21]

3.4.2 Discussion

The following conditions and events fostered the development of intramorphemic clitics in Udi: (i) the existence of a focus cleft, (ii) absence or loss of the copula, (iii) loss of the class agreement system inherited from Proto-Lezgian (see below), (iv) development of person-number clitics out of independent pronouns, (v) univerbation, (vi) a greater number of complex verbs than simplex (or loss of many simplex verbs), and (vii) analogical placement of PMs in simplex verb stems. Although it is impossible to say with certainty which of these factors are essential for the development of endoclitics, it appears that at least univerbation and analogical placement are crucial to this development. This last change, taken narrowly, is very unusual and thus is unlike the other, common changes described in this paper. But we may assume that analogical change can accomplish a great many things, and in the broad sense analogical change is very ordinary. Thus, each of these circumstances or changes appears to be common; it is the combination that is uncommon.

Perhaps factors (i–ii) above are not essential to the development of intermorphemic clitics; clitics can be trapped between parts of a verb without focus clefts. Factors (iii–iv) relate to the nature of the clitic, its agreement function. There are, of course, many other kinds of clitics in other languages, and in principle they could undergo all of these changes. It seems unlikely, however, that the analogical placement of clitics inside roots would develop with clitics that were not obligatory with every finite verb form, and in this sense, it seems most likely to occur with pronouns that are on the way

[20] The specific circumstances under which endoclitics fail to occur in Udi are detailed in Harris (2002a: 116–122).

[21] The story is more complex than this suggests; see Harris (2002a: chs. 9 and 11) for full details.

to becoming agreement marking. While changes other than univerbation, (v), might serve to get a morpheme inside a word, replacing one change with another is not simpler. Of course, we cannot be certain that (vi) is strictly necessary, but it seems likely that there would be insufficient pressure without this circumstance. Thus, although we cannot know for certain, it seems likely that at least the four changes or circumstances (iii–vii), or comparable alternatives, are essential to the development of intramorphemic clitics. Schulze (2004: 437) proposes an alternative to the hypothesized cleft in pre-Udi, but the development he outlines is not significantly simpler than that described here.

Some would argue that other languages have undergone univerbation, and none is known to have kept the structure at issue here. They might say that although change does explain how the unusual construction developed and why it was possible here and not in many other languages, it does **not** explain why Udi has chosen not to "repair" this construction, even though a way of doing so is straightforward. But there is nothing here in need of repair; it is self-evident that the system functions and can be acquired, since it has been used for at least 1,600 years (see note 18). Compare the aftermath of univerbation in Udi and univerbation in the Indo-European languages discussed by Watkins (1963, 1964) and by Jeffers and Zwicky (1980) and others. Udi kept and extended the construction, the others got rid of it; what is the difference? In both sets of languages, intermorphemic clitics developed in some verbs, while other verbs lacked them. There are at least three relevant differences between Udi and the IE languages that underwent univerbation. Any one of these, or any combination, may have been responsible for the different history of Udi. (i) In Udi, the overwhelming majority of verbs were new complex formations with the trapped intermorphemic clitic in some constructions, while in other languages there may have been a smaller percentage of verbs that had this construction. (ii) In Udi the inherited system of class–number agreement had recently been lost, was in the process of being lost, or was soon to be lost. The trapped morpheme provided a different kind of reference that could eventually be reanalyzed as agreement. This was apparently not the case in the IE languages. (iii) In Udi there is some reason to believe that some of the small set of simplex verbs had an existing "slot", a position held by the soon-to-be-lost class agreement affix or by one of the other morphemes that occurs intramorphemically in Udi (see Harris 2002a: 213–215, 219–221); in the IE languages discussed by Watkins, there is no reason to believe that such a "slot" existed.

At this point the skeptic might appeal to analogy as a powerful force in languages. Faced with a minority of verb forms (types and tokens alike) having intermorphemic clitics, most languages have regularized by losing this construction. Udi, on the other hand, was most likely faced with a vast majority of verbs (types) with intermorphemic clitics in some forms; regularizing across verbs took Udi in the opposite direction. However, this leaves us with the question of why Udi did not regularize all forms of a

given verb; if regularity is so important, why does =*ne* 'third-person singular subject' occur in one position in (17) and in a different position in (18)?

(17) ai-ne-z-o (*aiz-o-ne)[22]
arise$_1$-3SG-arise$_2$-FUTI
'She will arise'

(18) aiz-al-le (*ai-ne-z-al)
arise-FUTII-3SG
'She will arise'

(The tense-aspect-mood category in (18) is one of four that requires that PMs be enclitic to the verb form, while that in (17) makes no such requirement.) Udi has not regularized this particular difference for the same reason that other languages do not regularize everything. No language is entirely regular, and Udi is no different from others in this respect. Why does Italian not use one auxiliary with both unaccusative and unergative verbs? Why two? Regularization is common in languages, but there is no imperative to simplify.

Thus, I have proposed in this section that the development of the typologically unusual feature of endoclitics was possible in Udi because of the existence or occurrence of a particular set of conditions and events. Each of these circumstances or processes is familiar from other languages; what is unusual about them is their combination and perhaps their order. Udi has developed an unusual characteristic because of an unusual combination of common conditions and events. It is the product of historical accident, just as biological evolution is.

3.5 THE UNIFORMITARIAN HYPOTHESIS AND EXPLANATION OF TYPOLOGICALLY UNUSUAL STRUCTURES

The reconstructions that underlie the two explanations discussed here in detail—reconstructions of Common Kartvelian and of Proto-Lezgian—posit structures that appear to be simpler in relevant respects than their reflexes in the attested daughters. For Common Kartvelian I have reconstructed a single series, while all of the daughter languages, including Old Georgian, attest at least three series. Further, I have reconstructed a stage at which antipassives did not exist, though I have posited a later stage at which they did operate. For Proto-Lezgian I have reconstructed a stage at which there were no PMs, although Udi attests PMs. This does not violate the Uniformitarian Hypothesis, as I show below.

[22] This may be grammatical for some speakers but is rejected by my consultants. For details, see Harris (2002a: 136–138).

First, while I reconstruct Pre-Udi verb forms as lacking PMs, the system I reconstruct to Proto-Lezgian (see also Alekseev 1985; Schulze 1988, etc.) is by no means simple. While Proto-Lezgian lacked person markers, it had agreement through class markers (which Udi completely lacks except as relics, now part of the lexical stem of a few verbs), in something similar to the form in which daughters Rutul, C’axur, Budux, Kryz, and Arči have this today. Proto-Lezgian may have lacked the monoclausal focus construction found in Udi today, but is likely to have had focusing through clefts, just as many NEC languages have today (see Xajdakov 1986; Kazenin 1994, 1995, 2002, on focus clefts in NEC languages). The full complexity of independent pronouns found in Udi today and in some sister languages is also assumed for Pre-Udi (Harris 2002a: 179–183, 239). Thus, Pre-Udi and Proto-Lezgian (and Proto-NEC) were not, in fact, simpler than Udi; these languages were different but of comparable complexity in the relevant systems.

Common Kartvelian, in this respect, is quite different. While each daughter possesses at least three series of verb forms, I reconstruct one. While each daughter has, at least to some extent, an aspectual distinction between perfective and imperfective marked (in part) by preverbs, we reconstruct a system that lacked preverbs, though the distinction between perfective and imperfective is reconstructed as marked by series (see, for example, Kavtaraʒe 1954: 322; Schmidt 1966; Harris 1985: 98). Perhaps most disturbing, we reconstruct only four specific tense-aspect-mood paradigms (the precursors of the Old Georgian permansive (habitual), aorist, subjunctive, and imperative), although Old Georgian attests at least three times this many, and the sister languages attest comparable numbers. Nevertheless, no one suggests that CK was simple; for example, it had a complex system of ablaut (see Gamq’reliʒe and Mač’avariani 1965). Further, it seems likely that differences in third-person endings—singular *-s* vs. *-a* vs. *-n*, and plural *-en* vs. *-es* vs. *-ed*—may be relics of a tense, aspect, or mood opposition with complexity that cannot be fully reconstructed (Harris 1985: 98). With respect to tense-aspect-mood paradigms, while we can reconstruct only four specific paradigms, this does not entail that the protolanguage was limited to this number. But—and this is the important area of simplicity—it is indeed proposed that all of these verbal paradigms were in a single series, with a single case-marking pattern rather than the two or three found in each of the daughters. Note that while none of the daughters of the reconstructed language is limited to a single series, most languages of the world do make do with a single series. It seems that we are not out of line with the Uniformitarian Principle, inasmuch as the type we reconstruct is abundantly attested among languages of the world.

The Uniformitarian Hypothesis is often interpreted within linguistics as including the type of the language reconstructed; that is, it is often interpreted as the hypothesis that the types of languages found in ancient times[23] were not significantly different

[23] I understand “ancient times” here to refer only to the attested and reconstructible past; I assume that no version of the Uniformitarian Hypothesis is intended to include the origins of language.

from the types found in historical times. However, Comrie (2003) emphasizes in a recent essay that the original statement refers only to diachronic processes, not to states. That is, the Uniformitarian Hypothesis, as originally stated, provides that the diachronic processes that applied in earlier times are not different in kind from those that applied in historical times. As I have emphasized above, the processes that I and others have proposed to account for the development of complex case marking in Georgian and for the origin of endoclitics in Udi are of types known in other languages.

3.6 CONCLUSIONS

I have argued that two specific typologically unusual constructions in two languages of the Caucasus are due to the unusual co-occurrence of quite usual processes. The more changes are involved, the less likely all will happen to co-occur. Typologically unusual constructions can be explained in terms of their origin. This approach explains both the fact that such constructions are unusual and the fact that they occur at all. It is the fact that so many specific factors or changes must co-occur or occur sequentially in an appropriate order that explains the infrequency of these constructions, and no further explanation is needed. Many typologically unusual constructions can be explained as uncommon combinations of common changes. In this sense, they are the result of historical accident.

It is not surprising that probability would have a role to play in the origins of rare structures, and in that sense the results of this paper are not surprising. Nevertheless it is an idea that is necessary to discuss, because the role of probability has not been included in previous discussions of rare phenomena.

PART II

Phonological Universals: Variation, Change, and Structure

4

Consonant Epenthesis: Natural and Unnatural Histories

Juliette Blevins
Max Planck Institute for Evolutionary Anthropology

4.1 INTRODUCTION

Phonological rules of consonant epenthesis occur in many of the world's languages, and often involve insertion of a glide adjacent to a vowel. Phonological descriptions of consonant epenthesis tend to focus on the recurrence of prevocalic consonant insertion (1). Analysis of this recurrent sound pattern invokes two distinct types of universal markedness constraints. First, the position of the inserted segment is attributed to a syllabic markedness constraint which demands that syllables have onsets (2). Many researchers, including Jakobson (1929) and Greenberg (1978a: 75) suggest, on the basis of typological studies, that there is a strong preference for CV syllables, where C is a single consonant.[1] In a wide range of theoretical approaches (e.g., Prosodic Phonology, Optimality Theory, Government Phonology) rules of consonant epenthesis in intervocalic contexts are claimed to satisfy this preference, supplying the onsetless syllable with a requisite onset. A second component of the analysis attributes the quality of the inserted segment to segmental markedness constraints (3). Segmental markedness may be defined on the basis of articulatory parameters (e.g., Archangeli and Pulleyblank 1994), perceptual parameters (Steriade, in press), or some more abstract parameters (McCarthy 2002: 14–15), but the general claim is that

An early version of this paper was presented at the Explaining Linguistic Universals Workshop at the University of California, Berkeley in March 2003. Another version, focusing on Austronesian data, was presented at AFLA X at the University of Hawaii in March 2003. I am thankful to participants at both events for valuable feedback. Special thanks also to Andrew Garrett, Dave Kamholz, Ian Maddieson, Laurie Reid, and Andy Wedel for discussions leading to improvements in this version, and to Bob Blust for access to his Austronesian Comparative Dictionary, and other work in progress.

[1] Here and elsewhere, I use the term "consonant" in a non-technical sense to refer to non-vocalic non-syllabic segments including consonants, glides, and laryngeals. Phonetic symbols are those of the IPA, a period marks syllable boundaries. In phonological rule descriptions "Pr" stands for "prosodic", "PrWd" for "prosodic word", and "Cl-Pr" for a prosodic domain containing clitics.

universal properties of synchronic phonologies play a role in the determination of epenthetic consonant quality.

(1) A common phonological description of consonant epenthesis
There is a recurrent sound pattern $V > \mathbf{C}_i V$.

(2) Why the general pattern $\mathbf{V} > \mathbf{C}_i\mathbf{V}$? Universal syllabic markedness.
ONSET: *Syllables need / require onsets(* $*\ {}_{Syllable}[V\ldots]$*).*

(3) In $\mathbf{V} > \mathbf{C}_i\mathbf{V}$, why is C_i inserted instead of C_j? Universal segmental markedness.
C_i *is less marked than* C_j.

In this chapter, I suggest four general problems for the description in (1), and the proposed analyses in (2) and (3). First, many rules of consonant epenthesis (C-epenthesis) are restricted, at their origins, to either word-initial position, intervocalic position, or word-final position. In this case, the description in (1) is inaccurate. As I illustrate in section 4.2, a more accurate description of recurrent sound patterns involving C-epenthesis adjacent to vowels includes three distinct subcases, as spelled out in (4).

(4) Recurrent general patterns of consonant epenthesis (see section 4.2)

a. Intervocalic:	$\mathbf{VV} > \mathbf{VC}_i\mathbf{V}$
b. (i) Prosodic domain-initial:	${}_{\mathrm{PrWd}}[\mathbf{V} > [\mathbf{C}_j\mathbf{V}$
(ii) Prosodic domain-final:	$\mathbf{V}]_{\mathrm{PrWd}} > \mathbf{VC}_j]$

In the pattern described by (4a), an epenthetic consonant occurs between vowels, but not word-initially before vowels. In (4bi), an epenthetic consonant occurs at the beginning of a prosodic domain before vowels, but not intervocalically within the same domain. And in (4bii), an epenthetic consonant occurs at the end of a prosodic domain after vowels, but not intervocalically within the same domain. If, as I suggest below, these are the most common recurrent patterns of general consonant epenthesis adjacent to vowels, then the onset-filling account in (2) is difficult to maintain. Under (2), consonant epenthesis provides onsets for *all* onsetless syllables: the general pattern expected is one where consonant-insertion occurs word-initially before vowels *and* intervocalically. The rarity of such patterns severely undermines ONSET-based accounts.

Three other problems involve the segmental markedness constraints invoked in (3). First, segmental markedness constraints are unable to account for a striking cross-linguistic generalization: in the majority of cases where the historical phonology can be reconstructed, and where segments are *not* phonetically predictable, epenthetic consonants are precisely those for which earlier consonant loss is evidenced. In addition, segmental markedness constraints are unable to account for rare but attested cases of highly marked epenthetic consonants; in these cases also, historical rules of consonant loss are attested. A final problem for such accounts is that in some languages, the inserted epenthetic consonant is not a contrastive segment, and hence, is unlikely to be a direct consequence of phonological segmental markedness constraints.

I suggest that the four problems just noted arise from misguided analyses of consonant epenthesis grounded in universal synchronic constraints and markedness principles. Here, I present alternative diachronic explanations for a range of sound patterns involving consonant epenthesis, illustrating how and why reference to segmental and syllabic markedness should be replaced with substantive constraints on sound change, in the general spirit of Greenberg (1965, 1978). Like many other common sound patterns, regular consonant epenthesis may have a natural history, reflecting the phonologization of earlier phonetically conditioned sound change, or an unnatural history, reflecting something else. Unnatural histories include rule inversion, rule telescoping, where a sequence of natural changes yields cumulatively unnatural alternations, analogy, or language contact. The failure of universalist models to distinguish these diachronic origins results in the range of local problems for treatments of consonant epenthesis noted above, as well as non-local problems in phenomena in which the same set of universal constraints are claimed to play a role.[2] The study as a whole supports Greenberg's (1978a: 89) conclusions regarding the central nature of diachrony in defining and explaining cross-linguistic generalizations:

> Diachronic principles are involved in the explanation of both low and higher level synchronic generalizations. In so doing, they often explain exceptions. They also go even further than synchronic typology in subsuming under general principles not only non-universal typological traits, but often even highly idiosyncratic language-specific rules which can be treated as evidence of transitions between less complex, and more widely occurring types.

This study is framed within Evolutionary Phonology (Blevins 2004a, 2006a). Evolutionary Phonology shares with many contemporary approaches the view of language as a complex adaptive system where regular sound patterns are emergent probabilistic properties, resulting from the repeated interaction of innate perceptual and articulatory biases, self-organizing properties of sound systems, and aspects of language use within a population (e.g., Lindblom et al. 1984; Lindblom 1992; Steels 1997, 2000; Blevins and Garrett 1998; Bybee 1998, 2001, this volume; de Boer 1999, 2001; Pierrehumbert 2003; Mielke 2004; Wedel 2004; Oudeyer 2005). Within this framework, many recurrent sound patterns are argued to be a direct consequence of recurrent types of phonetically based sound change. Common phonological alternations like final obstruent devoicing, nasal-stop place assimilation, intervocalic consonant lenition, and unstressed vowel deletion, to name just a few, are shown to be the result of phonologization of well-documented articulatory and perceptual phonetic effects. Synchronic markedness constraints of structuralist, generativist, and optimality approaches are abandoned, and replaced, for the most part, with historical phonetic explanations which are independently necessary.

[2] Other models, like Natural Phonology (Stampe 1973), simply ignore sound patterns with unnatural histories.

The general Evolutionary Phonology approach, detailed in Blevins (2004a, 2006a), is summarized in (5) and (6).[3]

(5) Central premise of Evolutionary Phonology
Principled diachronic explanations for sound patterns have priority over competing synchronic explanations unless independent evidence demonstrates, beyond reasonable doubt, that a synchronic account is warranted.

(6) Hypotheses of Evolutionary Phonology supported by empirical investigation
a. Common sound patterns typically result from common phonetically motivated sound change.
b. Rare sound patterns are not the result of common phonetically motivated sound change.
c. Synchronic properties of particular sound patterns are better explained in diachronic terms than in terms of synchronic phonological universals.
d. Sound change is not goal-directed.
e. Rare sound patterns may be rare as a consequence of sound change, or may reflect accidental gaps in sound-pattern distribution.

Already, this framework has proven useful in identifying new phonetic explanations for well-documented recurrent sound patterns and for distinguishing sound patterns with a natural history in phonetic substance from those with an unnatural history involving rule inversion, rule telescoping, analogy, or language contact (Blevins 2004b, 2004c, 2004d, 2005a, 2005b, 2006a; Gessner and Hansson 2004; Hansson 2004; Mielke 2004; Yu 2004; Odden 2005; Shih 2005; Vaux and Samuels 2005; Garrett and Blevins in press; Iverson and Salmons 2005). Some general results of the model are summarized in (7). Specific results of this study fall into the same categories, and are summarized in (8). In this case, the general approach undermines the empirical truths expressed by (1)–(3), revealing, instead, those in (4). At the same time, the generalizations in (4) have diachronic phonetic explanations which need not be duplicated in the synchronic grammar.

(7) Some results of the Evolutionary approach
a. New common pathways of sound change are identified.
b. New phonetic explanations are proposed for previously problematic instances of sound change and sound patterns.
c. New non-phonetic explanations are proposed for recurrent sound patterns which defy phonetic explanation.
d. Markedness constraints are excised from synchronic grammars.

(8) Some specific results of this study
a. Epenthesis at the edge of prosodic domains is identified as a common sound change, having distinct properties from intervocalic glide epenthesis, as summarized in (4).

[3] For similar approaches to morphosyntactic regularities and rarities respectively, see Garrett (this volume) and Harris (this volume).

b. An association is found between laryngeal epenthesis and prosodic domains.
c. Regular epentheses which do not appear to be phonetically natural are, for the most part, the result of rule telescoping or rule inversion.
d. Explanations for general patterns of C-epenthesis adjacent to vowels do not require reference to syllabic markedness (2), or segmental markedness (3). Synchronic patterns of C-epenthesis provide no evidence for markedness constraints as components of synchronic grammars.

This chapter is organized as follows. Section 4.2 presents the two most common natural phonetic sources for epenthetic consonants adjacent to vowels: the reinterpretation of V-to-V transitions (4.2.1) and marking of prosodic boundaries with laryngeal features (4.2.2).[4] Section 4.3 presents the two most common unnatural sources for epenthetic consonants in intervocalic position: rule telescoping in which a once phonetically natural epenthetic segment undergoes fortition (4.3.1); and loss of weak consonants in coda position with rule inversion (4.3.2). Rarer cases of unnatural epentheses are illustrated in 4.3.3. Section 4.4 reviews evidence from general syllabification and prosodic morphology which also supports the elimination of syllabic and segmental markedness constraints from synchronic grammars. Section 4.5 presents a summary of findings as well as implications of this approach for capturing universal tendencies in sound patterns.

Before examining a range of consonant epenthesis rules, I should emphasize that by questioning the validity of ONSET (2) as a linguistic universal, I do not mean to question the important role of syllables in determining certain aspects of sound patterns. Arguments for the syllable as phonological constituent are summarized in Blevins (1995), and are not contested. What I do question is whether syllables and syllabifications are defined universally, by constraints like ONSET, or whether, as suggested by Greenberg (1978a: 75), Ohala and Kawasaki-Fukumori (1997), Steriade (1999), and Blevins (2003a, 2003c, 2004a, 2006b), syllable structure is, to a great extent, emergent, with syllabification arising as a consequence of the acquisition of word-based phonotactics.

4.2 NATURAL HISTORY

Synchronic sound patterns with natural histories are those which directly reflect the phonologization of earlier phonetically conditioned sound change. In this section I suggest that consonant insertion adjacent to vowels has two common natural histories each with a distinct set of distributional and phonetic properties. The first natural history involves the reinterpretation of a V-V transition as an intervocalic glide. The second involves the spontaneous occurrence of a non-contrastive laryngeal gesture at

[4] Epenthetic consonants which result from C-to-C transitions (e.g., ns > nts) or strengthening of release features (e.g., k^h > kx) are not discussed, since these give rise to new CC clusters, not CV syllables.

a prosodic boundary. Since nearly all regular sound change appears to be phonetically motivated, by identifying the most common natural histories for epenthetic consonants adjacent to vowels, we predict that C-epenthesis sound patterns which directly or indirectly reflect these sound changes will be more common than ones which do not. Recognizing these two distinct natural histories and the cluster of properties surrounding them goes a long way towards explaining why many of the world's languages have synchronic C-epentheses adjacent to vowels involving the glides *w*, *j*, or laryngeals *ʔ*, *h*.

4.2.1 Natural history I: the evolution of intervocalic glides

Glides often evolve spontaneously between adjacent vowels. There is little debate as to the phonetic explanation for this process. In hiatus contexts formant transitions between adjacent vowels can give rise to the percept of a medial glide. This historical process is most common when one of the vowels is /i/ or /u/, with reinterpretation of the vowel sequence as one with an intervening heterorganic glide: *ia* > *ija*, *ua* > *uwa*, etc. Examples of this phonetically natural sound change are illustrated in (9) from three different languages representing three distinct language families (Indo-European, Austronesian, and Papuan - Madang Adelbert Range subphylum).[5]

(9) Glide insertion as sound change

Language/source	Sound change	Examples			Gloss
a. Pre-Hindi	*ia > ija	*kia-	> ki**j**aa		'done'
b. Pre-Chamorro	*ua > uwa	*buaq	> *pu**w**a	> pu**gw**aʔ	'betel nut'
	*ia > ija	*liaŋ	> *li**j**aŋ	> li**dz**aŋ	'cave'
	*au > awu	*zauq	> *tʃawuq	> tʃa**g**oʔ	'far, distant'
c. Pre-Tauya	*ie > ije	*nie	> ni**j**e		'I/you (sg.) eat'
	*ia > ija	*nia	> ni**j**a		'they eat'
	*oe > owe	*oe	> o**w**e		'I/you (sg.) say'
	*ue > uwe	*-tue	> -tu**w**e		'I/you give to'

In languages like Chamorro the segmental status of earlier **w* is supported by the strengthening of **w> gʷ*, and **j > dz*. This strengthening effects directly inherited **w* and **j* (e.g., *gʷalu* < **walu* 'eight'; *adzudzu* < **qajuju* 'coconut crab') as well as predictable transitional glides like those in **puwa*, **lijaŋ* etc. At some point then, the phonetic glides had the same segmental status as those which were directly inherited.

Few linguists would deny the fact that sound changes like those shown in (9) exist, and that such sound changes have phonetic motivation. What needs to be stressed regarding examples of this type is that, first, no reference need be made to the featural make-up of the historically epenthetic glide: its quality is determined by the percept

[5] Pre-Hindi is from Hock (1991: 128). Pre-Chamorro is from Blust (2000). Tauya data is from MacDonald (1990), though in this case the historical reconstructions are my own, based on internal reconstruction.

arising from the transition from one vowel to the next. Since in this case, a principled phonetic diachronic explanation is evident, by (5), this historical explanation has priority over a synchronic markedness account which, at best, will only duplicate the diachronic explanation.

A second important point is that the historical evolution of transitional glides from ambiguous percepts is not altogether a chance event: the pre-existence of a glide *w* or *j* independent of the phonetic context in question appears to play a role in category acquisition in the majority of cases examined. In all languages listed in (9), contrastive *w* and *j* in other contexts predate the documented intervocalic sound change. This non-local effect on category acquisition is attributed to the general principle in (10). By the principle of Structural Analogy, a language learner is more likely to categorize ambiguous transitions as glides when pre-existing categories of glides exist in a language.

(10) Structural Analogy (Blevins 2004a: 247)
In the course of language acquisition, the existence of a (non-ambiguous) phonological contrast between A and B will result in more instances of sound change involving shifts of ambiguous elements to A or B than if no contrast between A and B existed.

The interpretation of ambiguous vocalic transitions as glides [j], [w] is predicted to be more common in languages in which the phonological categories /j/ and /w/ pre-exist than in cases where they do not. For example, in Pre-Chamorro, forms like **wada*, **walu*, **qawa*, **qajuju*, **daja*, **lajaR*, etc. were directly inherited with initial and medial contrastive glides. By (10), the acquisition of this contrast made it more likely for language learners to categorize phonetically predictable transitional glides like those in **buaq*, **liaŋ*, etc. as instances of /w/ and /j/ than if these categories were absent. In Rennellese, a Polynesian language without /w/ and /j/, the principle in (10) suggests that despite common *ia* and *ua* sequences, glide insertion as sound change is less likely to occur than in Pre-Chamorro since there are no pre-existing contrastive glides *w*, *j*, to serve as the basis of categorical analogy during the course of language acquisition.

A third point worth stressing for developments like those in (9) is that they do not support an ONSET constraint like that stated in (2): a requirement that syllables have onsets seems not only unnecessary, but also inaccurate. This stems from three factors: not all vowel sequences give rise to epenthetic glides; vowel-initial words may persist diachronically; and new vowel-initial words may arise as a consequence of subsequent initial C-loss and/or borrowing. Each of these points is illustrated in (11) for Lou (Austronesian), a Western Admiralties language with transitional glides in phonetically predictable contexts. Data in (11) is from Blust (1998). In (11a) we see that not all vowel sequences in Lou give rise to intervocalic glides; in (11b), inherited vowel-initial words are maintained without change. And in (11c), new vowel-initial words have evolved as a consequence of initial C-loss and borrowing, again, with no obvious constraint

against vowel-initial syllables which one might expect if a constraint requiring onsets were at work.[6]

(11) Lou intervocalic glide-insertion ≠ Onset

a. Not all vowel sequences give rise to intervocalic glides (only sequences of rising sonority do)

	/tia-n/	[tijan]	'his/her abdomen'
	/kea/	[keja]	'swim'
	/moloa-n/	[molowan]	'his/her shadow/spirit'
	/suep/	[suwep]	'digging stick'
But:	/wei-n golom/	[weiŋgolom]	'your saliva'
	/mween/	[mwɛɛn]	'man, male'
	/kapeun/	[kaβeun]	'bitter'

b. Vowel-initial words are typically unaffected phrase-initially

/okok/	[okok]	'to float'
/ara-mu ŋata-n/	[arɔmŋaran]	'your head is bald'
/i lɪp nɔt/	[ilɪpnɔt]	'pregnant', (lit. 'she is carrying a child')

c. New vowel-initial words may arise via initial C-loss or borrowing

i. *p > Ø / _V (Pre-Lou) **paŋan > aŋ* 'feed'; **pia > ia-n* 'good', **puka > uk* 'open, uncover', **apaRat > aa* 'south wind', **sa-ŋapuluq > saŋaul* 'ten'.

ii. *k > Ø /_V (Pre-Lou) **ka > a* 'and', **keri > er* 'scrape out', **konso > os* 'husk coconuts'.

iii. *aipika* 'an edible plant: *Hibiscus manihot*', (loan: NG Pidgin *aipika*)

One final argument against the syllable onset as a catalyst for sound changes like those in (9) is the fact that parallel phonetic developments occur in other positions within the syllable. As shown in (12), homorganic glides can evolve from reanalysis of complex VC transitions giving rise to new complex nuclei (12a), or coda glides (12b), while the reinterpretation of complex CV transitions can yield glides as components of complex onsets (12c).

(12) Homorganic glide/vowel evolution in non-onset position (Hock 1991: 119–120)

Language	Sound change	Examples	Gloss	Output of sound change
a. American English	ʃ> jʃ, ʒ > jʒ	mæʃ> mæjʃ	'mash'	coda, complex coda
		meʒɹ̩ > mejʒɹ̩	'measure'	
b. Old French	ɲ> jn > in	*plaɲit > *plaint*	'complains'	complex nucleus
		*poɲu- > *poing*	'fist'	
c. Lithuanian	p^j > pj	*p^jaut^{j}i > pjauti	'cut'	complex onset

What the examples in (9) and (12) share is the perceptual ambiguity of formant transitions between historically adjacent segments. In all examples, the formant transitions

[6] Since the changes in (11ci,ii) arguably precede glide insertion historically, they only show that no glide has been accreted word-initially.

Another sound change which directly eliminates onsets is glide-vocalization. Such a change appears to have occurred in Muna, an Austronesian language of Muna island: PAN **walu* 'eight', Muna *oalu*, PCMP **waiR*, Muna *oe* (van den Berg 1989).

are interpretable as segmental glides and the phonetic quality of the glide is predictable. However, unlike the sound changes in (9) which may be interpreted as ameliorating syllable structure, those in (12) result in more complex syllable types.

In sum, the evolution of glides *w* and *j* as reinterpretations of formant transitions in relevant V-V, C-V, and V-C strings is part of the natural history of consonant epenthesis. In these cases, the two most important explanatory factors appear to be: (i) the occurrence of formant structure which is perceptually similar to that of a glide; (ii) the occurrence of unambiguous glides outside of the context in question for pattern matching in the course of language acquisition in accordance with (10).

4.2.2 Natural history II: laryngeals at prosodic boundaries

While segmental glides *j* and *w* often evolve spontaneously between adjacent vowels, epenthetic laryngeals *h* and *ʔ* have an arguably different natural history resulting in distinct sound patterns and system-internal relationships. Significant differences are summarized in (13). One important difference concerns distribution: where epenthetic segmental glides *w* and *j* typically originate in intervocalic contexts, regular sound change involving epenthetic laryngeals typically originates at prosodic boundaries, including the phonological phrase and phonological word. Another difference between the two types of sound change is their contrastive status. Sound changes like those in (9) are most common within sound systems in which contrastive glides are pre-existing; however, the occurrence of a laryngeal closing or spreading gesture at a prosodic boundary is typically non-contrastive at its point of origin, and may continue to be non-contrastive for many generations.

(13) Significant differences between *j/w* and *ʔ/h* epenthesis-as-sound-change

	j/w	*ʔ/h*	Languages with prosodic epenthesis
Position of origin	V_V	$_{Pr}$[_V	English (ʔ), French(ʔ), Rennellese (ʔ), Ritwan (h), Nhanda (h/ʔ), Muna (ʔ), Anejom (ʔ)
		V_]$_{Pr}$	Chintang (ʔ), Yurok (h), Atayal, etc. (ʔ); Aklanon, etc. (h) [see (15)]
Typically contrastive	Yes	No	
Input phonetic ambiguity	Yes	not necessarily	

English presents a typical example of prosodic laryngeal distribution. In English, vowel-initial words are preceded by glottal stop at the beginning of the phonological phrase or word. Glottal stop, however, is not contrastive, and in fact, many English speakers have great difficulty perceiving glottal stop word-initially, or producing vowel-initial words not preceded by glottal closure. If this segment is not *perceptible* to the average English speaker, it is quite odd to refer to it as satisfying a constraint which requires that syllables have onsets. At the same time, within the phonology of English where contrasts are encoded, there is no contrast between words beginning in *ʔV*... and those beginning in *V*....

Similar facts are reported for many languages around the world. For example, in Muna, word-initial vowels are optionally preceded by a non-phonemic glottal stop, as illustrated in (14a) (van den Berg 1989: 21–27). In Anejom (a.k.a. Aneityum, Lynch 2000), utterance-initial glottal stop precedes vowels, and again, is non-contrastive. Compare the utterance-initial vowel in (14b) with its non-utterance-initial counterpart. Another language with similar sound patterns is Rennellese (Elbert 1988: 7). Although glottal stop contrasts with both zero and other consonants in word-medial position, "its distribution differs from that of other consonants in utterance-initial position" where "it occurs predictably before words that otherwise begin with vowels .... This predictable glottal stop may be considered a feature of initial vowels."

(14) Non-contrastive phrase-initial [ʔ] in Muna, Anejom, and Rennellese

	i. Prosodic laryngeal; Phrase-initial	ii. No prosodic laryngeal; V-initial syllables	Gloss
a. Muna	[ʔina]	[i.na]	'mother'
	[ʔo.e]	[o.e]	'water'
	[ʔurɛ]	[u.rɛ]	'high tide'
		[a.i.ni]	'this'
		[wa.e.a]	'bat'
b. Anejom	[ʔa.ek] 'you (sg.)'	[ʔa.ba.ma.ek]	'you (sg.) come!'
c. Rennellese	[ʔe uʔu e te hokai]		'the lizard bites'
	[ʔuʔu mai]		'bite!'

Consider the pronunciation of the verb 'bite' in the two phrases in (14c). Where the verb *uʔu* is not phrase-initial, there is a phonetic transition between vowels *e-u*, with no glottal gesture; where the same verb is utterance-initial, it is preceded by glottal stop. Though the medial glottal stop in *uʔu* 'lizard' contrasts with *u*: and other *uCu* sequences, glottal stop does not contrast with zero word-initially. There is no evidence for a general onset requirement in Muna, Anejom, or Rennellese. In each language, accounting for word- or phrase-initial glottal stop by invoking ONSET will require stipulations as to why the same constraint does not apply to every medial vowel-initial syllable.

A further argument against laryngeal epenthesis as serving a basic onset-filling function is the fact that it is not limited to the beginning of a prosodic domain. As shown in (13), edge-final prosodic laryngeals are also found and are well described for many Austronesian languages (Blust, to appear: 25). As summarized in (15), these sound patterns include word-final non-contrastive glottal stop in Formosan languages, Bashiic languages, languages of Borneo, and Sundanese of West Java (15a), as well as final [h] insertion for a range of central Philippine languages, and for languages of northern and central Sarawak (15b).[7]

[7] Given the number of independent instances of final glottal-stop epenthesis, Blust (to appear: 25) suggests the possibility that "final glottal stop was a feature of word endings in PAN." The symmetry of laryngeal epenthesis at opposite ends of the prosodic domain, along with the cross-linguistic frequency

(15) Final laryngeal epenthesis in Austronesian (Blust, to appear)

Rule	Languages
a. Ø → ʔ/ V_]$_{PrWd}$	Atayal, Saisiyat, Pazeh, Bunun, Kavalan, Paiwan, Puyuma, Amis; Yami, Itbayaten, Ivatan, Casiguran Dumagat, Brunei Malay, Sarawak Malay, Taboyan, Lawangan, Kapuas, Ba'aman, Katingan, Dohoi, Murung, Tunjung, Lou, Sundanese.
b. Ø → h / V_]$_{PrWd}$	Aklanon (and other Bisayan dialects of C. Philippines), Tagabili, Taosug; many northern and central Sarawak lgs., including: Miri, Narum, Kiput, Berawan, Western Penan, Long Wat Kenya, Sebop, Kelabit, Dalat, Matu, and Serike Malanau.

Let us look at one example which highlights the relationship between laryngeals and prosodic boundaries. In Aceh (Durie 1985: 36–37), an Achinese-Chamic language, *h*-epenthesis is described for a very particular prosodic boundary: "Clitics which have no syllable-final consonant add /h/ when they are enclitics, and occur last in the phrase." Compare the examples in (16a) and (16b): in the first set, a clitic (*neu, pi*) is phrase-final and occurs with epenthetic [h]; in the second set the clitic (*ka, pi*) is not phrase-final and no [h] is inserted. Notice that in these examples, onset plays no role, since the inserted [h] is in coda position.

(16) Final [h]-epenthesis in Aceh (Durie 1985) IN = inchoative, EMPH = emphatic
Ø → h / V_]$_{Cl\text{-}Pr}$

a. droe=neu**h**	ka=neu=jak	b. ka=droe-neu-jak	
self=2	IN=2=go	IN=self-2-go	
'you have gone'		'you have gone'	
lôn=pi**h**	sakêt	peulandôk	pi=ji-beudöh
I=EMPH	sick	mousedeer	EMPH=3-rise
'I am sick too!'		'The mousedeer got up.'	

Another language which illustrates an association between laryngeal settings and prosodic boundaries is Yurok. In Yurok and Wiyot, the two Ritwan languages within the Algic family, historically vowel-initial words have acquired an initial *h*. Comparison sets are shown in (17), with the proposed sound change in (18).

(17) Word-initial *h* in Wiyot and Yurok : Proto-Algonquian vowel-initial words
a. PA **e:wa* 'he goes' < ***a:*- 'go'; Y **ho*- in *heɣok'* 'I go' (intensive infix -*eɣ*-, 1SG -*k'*); W **ho*- in *hol*- 'go, walk'.
b. PA **ekwa* 'the other says so to him' (A294); cf. Y *h*- in *hek'* 'I say', *hiʔ* 'it is said', W *h*- 'say to', *hi*- 'be said'.

(18) Ritwan sound change: Ø > h / $_{PrWd}$[_ V

Some evidence for the prosodic conditioning of this rule is the sandhi alternations which are continued in both Yurok and Wiyot (Blevins and Garrett 2007). First, in

of such patterns in non-Austronesian languages, and the variability within Austronesian of final [h] vs. [ʔ] are all consistent with phonetic multigenesis.

both languages, there are *h*-initial nouns which surface without *h* under pronominal prefixation. An example from Yurok is shown in (19a). Second, in both languages, word-initial /h/ alternates with [j] after /i/, suggesting that in *i#V* sandhi contexts, where V was not initial in the prosodic word, there was no [h]. A representative example is given in (19b).

(19) Yurok initial /h/ in sandhi contexts

a.	Word-internal:	haʔaaɣ	'rock'	cf.	nepuj	'salmon'
		'naʔaaɣ	'my/our rock'		'ne-nepuj	'my/our salmon'
	h∼ Ø	'waʔaaɣ	'his/her/its/their rock'		'we-nepuj	'his/her/its/their salmon'
		k'a'aaɣ	'your rock'		k'e-nepuj	'your salmon'
		pelin haʔaaɣ	'big rock'		pelin nepuj	'big salmon'
b.	External:	heɣo'l	'he goes'			
		ni [j]eɣo'l	'he goes there'			
	h∼j	hunowoni	'growing'			
		k'i [j]unowoni	'the growing (things)'			

Note that the sound change in (18) is limited to prosodic boundaries, and therefore cannot be attributed to a general syllable onset requirement. The internal sandhi facts in (19a) suggest vowel cluster reduction, while the external sandhi facts in (19b) suggest historical Ø > j / i_V.

The historical process in which *h* was inserted at the beginning of prosodic words in Ritwan is mirrored word-finally in Yurok where there is evidence for *h*-insertion at the *end* of prosodic domains. A striking aspect of Yurok sound patterns is that there are no nouns or verbs ending in short vowels, with the sole exception of attributives ending in *-ni*. Where a word-final short *V* is expected, final *Vh* surfaces instead. For example, where vowel shortening has occurred in the *o*-class paradigm for *new-* 'to see', we expect the first plural indicative *newoo* to shorten to *newo*, but the surface form is *newoh*. As should be clear from the words in (20), syllables ending in short vowels appear freely in non-final position.[8]

(20) Word-final syllable types in Yurok nouns and verbs

Rhyme	Noun	Gloss	Verb	Gloss
VV	te.poo	'fir tree'	roo	'to be a particular time'
VC	le.wet	'salmon net'	ne.pek'	'I eat'
Vʔ	hoʔ.mo.noʔ	'tan oak'	he.ɣiʔ	'it was said'
Vh	ho.mo.nah	'live oak'	ne.poh	'we eat'
	pi.ʔih	'mussel'	ne.woo∼ne.woh	'we see'
VVC	ha.ʔaaɣ	'rock'	ne.wok'∼ne.wook'	'I see'
VVʔ	pe.co.looʔ	'ko sugar pine'	ske.wi.paaʔ	'he put in order'
V	—		—	

[8] See Berman (1981) for a discussion of laryngeal increments before voiceless stops and Blevins (2003b) for more on Yurok syllable structure.

In order to account for the absence of word-final short *V*, and instances where expected word-final V surfaces with final *Vh*, the sound change in (21) is proposed.

(21) Yurok sound change: $\emptyset > \mathrm{h}\ /\ \mathrm{V}_]_{\mathrm{PrWd}}$

Since (21) applies at the end of prosodic words and phrases, we expect to see sandhi alternations similar to those seen for word-initial /h/. This is the case. As shown in (22), locative suffixation suggests earlier noun-forms lacking final /h/.

(22) Yurok final /h/ in sandhi contexts (*-oɬ* 'locative')

a. Word-internal:	paʔah	'water'	cf.	nepuj	'salmon'
	paʔaaɬ	'water.LOC'		nepujoɬ	'salmon.LOC'
h∼ Ø					
	tektoh	'log'		looɣin	'fish.dam'
	tektooɬ	'log.LOC'		looɣinoɬ	'fish.dam.LOC'
b. Word-internal	piʔih	'mussel'			
h∼ j	piʔijoɬ	'mussel.LOC'			

The association of non-contrastive laryngeals with prosodic boundaries is a recurrent sound pattern in languages across the world. But why is it natural for laryngeals [h] and [ʔ] to mark prosodic boundaries, and why are these segments typically non-contrastive at origin in these positions? A preliminary answer relates the occurrence of default laryngeals to pitch contours which characterize prosodic boundaries. Prosodic boundaries are typically marked by pitch contours initiated via laryngeal mechanisms. It is this laryngeal involvement which likely gives rise to fixed articulatory laryngeal gestures at certain prosodic boundaries.[9] While this is a preliminary hypothesis, the high frequency of such sound patterns, their regularity, and their subphonemic nature all point to natural phonetically conditioned origins.

Intervocalic glide epenthesis and prosodic laryngeal epenthesis constitute the majority of phonological epentheses where a synchronic alternation may have traceable roots to natural phonetically conditioned sound change (*pace* note 4). The segmental quality of an intervocalic glide need not be specified, since its quality follows from the phonetic characteristics of surrounding vowels. In laryngeal epenthesis, a glottal spreading or closure gesture may come to be specified as an aspect of language-specific phonetics precisely where this gesture is non-contrastive. Reference to onset in both types of epenthesis does not appear justified: in intervocalic glide epenthesis, word- or utterance-initial vowels are typically unaffected, while, under prosodic laryngeal epenthesis, domain-medial vowel-initial syllables are unaffected. In particular cases like Aceh or Yurok domain-final *h*-epenthesis, there are no cases where onsets are created, and prosodic domains must be invoked. Having examined C-epenthesis sound patterns with natural histories, I now turn to unnatural histories, where rule inversion, rule telescoping, analogy, or language contact may be involved.

[9] In at least one case where pitch accent is involved, in the Okinawa dialect of Japanese, the loss of distinctive pitch has given rise to a final glottal stop (John Whitman p.c., 2003).

4.3 UNNATURAL HISTORY

Synchronic sound patterns with unnatural histories are those which do not directly reflect the phonologization of earlier phonetically conditioned sound change. In this section I suggest that intervocalic consonant insertion has two fairly common unnatural histories each with a distinct set of distributional and phonetic properties. The first unnatural history involves two sound changes in sequence: a natural sound change involving intervocalic glide insertion or insertion of a laryngeal, followed by subsequent strengthening of the glide or laryngeal. As a consequence of this sequence of sound changes, phonetically unnatural rules of intervocalic obstruent epenthesis may be in evidence in synchronic grammars. The second common unnatural history involves the inversion of an earlier rule of consonant loss. Recognizing these two distinct unnatural histories and the cluster of properties surrounding them goes a long way towards explaining why many of the world's languages have synchronic C-epentheses involving segments other than the glides *w*, *j*, or laryngeals *ʔ*, *h*, and accounts for many of the distributional properties of these segments.

4.3.1 Unnatural history I: epenthesis + strengthening

Intervocalic glides in many languages are susceptible to further sound changes. As illustrated earlier in (9b), Chamorro glides underwent context-free strengthening, with **w* > *gw* and **j* > *dz*. The combination of phonetic glide epenthesis with subsequent glide strengthening constitutes a common instance of rule telescoping. If the epenthesis alternations are maintained synchronically, an unnatural sound pattern is in evidence as a consequence of sequential natural sound changes. The alternations in (23) from Chamorro illustrate a case in point, where (23b) shows intervocalic insertion of [dz] between a vowel-final verb stem, and /-i/ the referential focus suffix, and (23c) shows strengthening of /w/ to [gw] in the same context. These alternations are the result of the sequence of sound changes in (23i, ii) (Blust 2000).

(23) Chamorro obstruent~zero alternations from glide-epenthesis + strengthening
Sound change i. Vi > Vji, Vu > Vwu
Sound change ii. j > dz, w > gw

a.	a.mo.ti	'take away for'	cf. amot	'take away'
b.	ha.tsa.dzi	'lift for'	cf. ha.tsa	'lift'
c.	ha.na.gwi	'go for'	cf. ha.naw	'go'

(24) Singhi obstruent/zero developments (PMP = Proto-Malayopolynesian)
from laryngeal epenthesis + strengthening (Blust, to appear)
Sound change i. Ø > h / V_$]_{\text{PrWd}}$
Sound change ii. h > x/ u__
s/ i__

a. PMP *Raja > aju**x** 'great, large'
b. PMP *batu > batu**x** 'stone'
c. PMP *qubi > bi**s** 'yam'
d. PMP *kali > kari**s** 'dig'

Epenthetic laryngeals like those illustrated in 4.2.2 may also become phonologized, and are also susceptible to further sound changes. In Singhi (a.k.a. Land Dayak), as illustrated in (24ii), insertion of a word-final laryngeal is followed by context-sensitive strengthening conditioned by vowel context: coarticulation of *h* with a high back vowel yields *x*, while coarticulation with a high front vowel yields *s*. In Singhi, no synchronic alternations are involved, since reflexes of word-final **h* remain word-final.

However, in cases where a laryngeal/zero alternation exists, as seen in (19) for Yurok, subsequent changes to the laryngeal can give rise to unusual alternations. In Yurok, intervocalic /h/ is realized as ɣ, a voiced velar fricative. As a consequence, historically vowel-initial stems are ɣ-initial when preceded by vowel-final particles within the phonological word. Since the *h*/ɣ alternation remains transparent in Yurok, there is synchronic evidence for both *h*-insertion/deletion and *h*-strengthening, as illustrated in (25). However, the general point is that seemingly unnatural surface Ø/ɣ alternations exist as a consequence of an intermediate stage of *h*-strengthening.[10]

(25) Yurok initial /h/ in sandhi (Robins 1958)

	haʔaaɣ	'rock'	cf.	nepuj	'salmon'
h∼ Ø	'naʔaaɣ	'my/our rock'		'ne-nepuj	'my/our salmon'
	'waʔaaɣ	'his/her/its/their rock'		'we-nepuj	'his/her/its/their salmon'
	k'aʔaaɣ	'your rock'		k'e-nepuj	'your salmon'
h∼ ɣ	ku ɣaʔaaɣ	'the rock'		ku nepuj	'the salmon'
	ku $_{PrWd}$[haʔaaɣ	'the rock'			

4.3.2 Unnatural history II: coda loss + rule inversion

One of the most striking observations regarding regular consonant epenthesis is that once one leaves the domain of phonetically predictable glides and prosodically conditioned laryngeals, the range of epenthetic consonants is, for the most part, coextensive with consonants which are most commonly subject to weakening and loss in syllable-coda position. While this collection of segments includes glides *w/j* and laryngeals *h*/*ʔ*, it also includes liquids and nasals which are unattested as direct outputs of the phonetically natural processes examined in section 4.2. A further difference relates to contexts of consonant insertion. Where natural epentheses allow statement of the conditioning environment in terms of positive intervocalic contexts, some unnatural epentheses occur in intervocalic contexts which appear to be the complement to specific postvocalic coda environments. In (26), differences between the natural

[10] Notice the similarity between the Yurok change and that in Singhi. In both languages a strengthened **h* is realized as a velar fricative. In Yurok, regular strengthening is limited to intervocalic position where voicing also occurs.

epentheses illustrated in section 4.2 and common consonant-zero alternations which do not directly reflect such sound changes are summarized.

(26) Significant differences between synchronic epentheses with natural and unnatural histories

	Natural history	Unnatural history
Position	V_V	complement of C > Ø / V_i _.
	$_{Pr}$[_V, V_]$_{Pr}$	
Segment quality	w/j, h/ʔ	w/j, h/ʔ, **nasal, liquids**

The differences just noted are clearly not accidental: numerous instances of phonological C-epenthesis between vowels are clearly instances of historical rule inversion where an earlier sound change of postvocalic consonant loss is reanalyzed as consonant insertion (Vennemann 1972b; Blevins 1997; Vaux 2002). As a consequence, inserted consonants are precisely those which have been subject to historical weakening and loss, and the position of epenthesis is typically the complement environment to that of original loss. Rule inversion, as illustrated in (27), is the consequence of an analytic problem in the course of language acquisition. Having established a systematic pattern of alternation, the learner is faced with the question of which alternate is basic and which is derived. If the non-historical alternate is taken as basic, rule inversion may occur.

(27) Rule inversion resulting in C-epenthesis, where W = weak consonant
a. Sound change: W > Ø / V_i_ {#, C}
b. Resulting surface patterns: V_i{#, C} ~ V_iWV
c. Synchronic reanalysis: Ø → W / V_i_ V

A general prediction of the current model is that a consonant which undergoes natural phonetic weakening/loss in postvocalic or intervocalic position can be reanalyzed as epenthetic in precisely the contexts where it was *not* lost, since surface patterns of consonant~zero alternations typically give rise to phonological ambiguity: is the predictable consonant underlying, or inserted by rule? If a language learner chooses to derive predictable consonants by rule (27c), rule inversion can occur, resulting in a wider range of surface epenthetic consonants than those resulting from phonetically natural processes.

In (28) I list a range of languages where coda loss of a weak consonant has been inverted, giving rise to epenthesis of the consonant which was historically lost.[11] Blevins (1997) illustrates the extent to which C-epentheses of the sort illustrated in (28) are problematic for segmental and syllabic markedness accounts. While these accounts can handle the facts, they are forced to treat laterals, uvular fricatives, or labialized pharyngealized rhotic glides as unmarked segment types.

[11] For other potential examples and some implications for synchronic grammars, see Vaux (2002).

(28) Examples of inverted C-loss giving rise to C-epenthesis

Language	Consonant	Epenthesis context	
Uradhi	ŋ	V#_V	(Hale 1976)
English, RP, Boston, etc.	ɹ	V_i#_V (V_i= lax)	(refs. in Blevins 1997
Bristol, Mid.At. etc.	l	V_i#_V (V_i=lax; V_i =ɔ)	and Gick 1999)
Yupik	ʁ	V_i_-V ($V_i \neq$ ə)	(Jacobson 1984)
Anejom	ɾ	V_ #V	(Lynch 2000: 29)

Furthermore, like the instances of intervocalic glide epenthesis reviewed earlier, accounts making reference to ONSET fail to explain why C-epenthesis is limited to intervocalic contexts and does not insert the same unmarked consonant before vowels in phrase-initial position.[12]

Synchronic consonant~zero alternations are also attested in cases where a variety of consonant types underwent loss in a particular position. Perhaps the most well-studied case of this type is the loss of final consonants in the history of the Oceanic languages. Proto-Oceanic had many intransitive verbs and nouns which were consonant final. However, several daughters of Proto-Oceanic regularly lost consonants in word-final position. As a consequence of the historical sound change in (29), an extreme case of (27a), many Oceanic languages have C~zero alternations, where a consonant of unpredictable quality surfaces in suffixed forms and is absent in unsuffixed forms.

(29) Final C-loss in Oceanic (Ross 1998)
Proto-Oceanic *C > Ø / $_]_{PrWd}$ in, e.g.: Central Pacific (Fijian, Rotuman, Polynesian) Western Oceanic (Manam, etc.) South-East Solomonic (Toqabaqita, etc.)

The consonants which surface only under suffixation are traditionally known as "thematic consonants". Since, historically, they constitute the set of inherited final consonants, they cover a range of obstruents and sonorants, and include consonants at multiple points of articulation. In (30) I give a list of Oceanic languages which show C~Ø alternations under suffixation, and the range of attested thematic consonants.[13]

[12] See Bermúdez-Otero and Börjars (2006) for a defense of markedness in the case of English intrusive [l] in one dialect. Their general claim, as I understand it, is that contexts of rule inversion are determined, not by the original phonetic context of the consonant loss in question, but by markedness constraints which set limits on grammatical restructuring.

[13] Thematic consonants before long and short suffixes are collapsed, and *-ina* < **-nia* is treated as *n*-initial. For further discussion of these alternations, see Hale (1973a), Lichtenberk (2001), Pawley (2001), and Blevins (2004b).

(30) C~Ø alternations in some Austronesian languages
(Hale 1973a, Lichtenberk 2001, Pawley 2001)

Language	Thematic Cs	Extension of thematic C?
Toqabaqita	t, f, s, m, n, ŋ, l, r, w, ʔ	no
Manam	t, m, n, ŋ, l, r, w, ʔ	yes, semantic (see 31c)
Maori	t, k, m, n, ŋ, r, wh, h	default (see 31b)
Samoan	t, s, f, m, n, ŋ, l, ʔ	no
Tongan	t(s), k, f, m, n, ŋ, h, ʔ	no
Niuean	t, k, m, n, h	no
Hawaiian	k, m, n, l, h, ʔ	yes, -*ʔia*

If segmental or syllabic markedness constraints played a role in synchrony or diachrony, we might expect: (i) a general shift from marked to unmarked consonants in these contexts; (ii) unmarked consonant insertion in cases where the lexical thematic consonant is not retrievable; (iii) extension of C-insertion to other V_V contexts. However, there is little evidence for any of these predictions, as summarized in (31).

(31) Some observations regarding C~Ø alternations in Oceanic
a. /f/, /m/ are maintained in many Oceanic languages in this context
b. Maori default passives are sensitive to prosodic structure of the base, and there are different 'default' consonants in different dialects (Hale 1973a; Blevins 1994)
i. -*ia*, -*a* when base is bimoraic
ii. elsewhere, -*Cia* dialect I: -*tia*
dialect II: -*hia*
dialect III: -*ŋia*
c. Where leveling does occur within subparadigms, it is inconsistent with phonological markedness predictions.
Manam (Lichtenberk 2001), transitive verbs based on kinship all take -*m*-, where this pattern is arguably an innovation.

tama-m-i	'regard s.o. as one's father'
tina-m-i	'regard s.o. as one's mother'
toqa-m-i	'regard s.o. as one's older same-sex sibling'
tari-m-i	'regard s.o. as one's younger same-sex sibling'
natu-m-i	'adopt s.o. as one's child'
taua-m-i	'have s.o. as one's trading partner'
ruaŋa-m-i	'have s.o. as one's friend'

d. V-initial suffix allomorphs are maintained without change.
In Maori, 41.09 percent of passives in Biggs (1966) take -*a*.

First, consider the suggestion that segmental markedness constraints might shape the nature of these alternations over time, leveling them to less marked consonants. In cases where alternations are productive or semi-productive, alternations between /f/ and /m/ are maintained in most Oceanic languages (31a). Under markedness accounts, this might be unexpected. In cases where the lexical consonant is not retrievable, there are at least three different strategies reported for Maori dialects alone.

First, the vowel-initial suffix is used when a base is bimoraic; when bigger than the bimoraic minimal word, forms take a C-initial suffix, with C variable across dialects (31b). One dialect may reflect type frequency, since *-tia* (31.1 percent of passives from Biggs 1966) has the highest frequency of any *-Cia* suffix. However, other dialects may reflect token frequency, since many common verbs occur with *-hia* and *-ŋia*. In careful studies where leveling within subparadigms has been examined, consonants appear to take on semantic domains, rather than undergoing phonological bleaching. For example, as shown in (31c), *-m-* has been extended in Manam within a subclass of semantically related transitive verbs. Since many will agree that a phonological approach to these alternations is not warranted, it is worth pointing out that the hiatus occurring between vowel-final stems and the historic vowel-initial suffix has not been altered by regular epenthesis in the majority of languages (31d).

Clearly the most general observation one can make with respect to alternations arising from Oceanic final consonant loss, or similarly general processes of final C-loss like that which occurred in the history of French, is that phonological rule inversion of the sort schematized in (27) simply does not take place: since consonant quality is not predictable, no phonological generalization is made, and alternations are gradually lexicalized, with emergent developments or generalizations for the most part governed by morphological or semantic analogy.[14]

In sum, as with the natural epentheses investigated in 4.2.2, there is no need to invoke segmental or syllabic markedness constraints to account for regular consonant epenthesis originating in historical rule inversion. The schema in (27) suggests that any consonant which can be lost through natural phonetic processes in the syllable coda can give rise to regular epenthetic alternations, independent of its purported segmental markedness. Furthermore, the historical approach maintains that, in general, such alternations will not be extended to initial vowels. This is the case for the alternations listed in (28), as well as others documented in Vaux (2002).

4.3.3 Other unnatural histories

In this section two instances of historical consonant epenthesis are presented which reflect uncommon developments due to the unlikely convergence of phonetic, phonological, morphological, and syntactic properties. Each case has as its point of origin a type of sound change already discussed. Oceanic *j*-accretion, in subsection (i), may originate from high frequency *ija* (< *i#a*) strings in these languages. Ritwan *l*-sandhi in subsection (ii), on the other hand, has its source in rule inversion of earlier **l*-loss, combined with phonetic glide insertion.

[14] For discussion of constraints on analogical change, see Garrett (this volume) and Albright (this volume), who discuss the same Maori developments.

(i) What is Oceanic j-accretion?

Blust (1990) describes a recurrent development in Oceanic languages which involves the word-initial insertion of a palatal glide *j* before /a/. In (32) I summarize the sound changes in question.

(32) Glide accretion in Oceanic (Blust 1990)

Language	Sound change	Remarks
Fijian	* Ø > j / $_{\text{PrWd}}$[_a	Not in Waya, Nakoroboya (West); not in Labasa, Dogotuki (NE Vanua Levu); in Bua, optional or lexically determined. Lexically gradual change (*j > θ intervenes) /j/ has low functional load. (Geraghty 1983)
Gedaged	* Ø > j / $_{\text{PrWd}}$[_a	Milke (1968) suggests this sound change is widespread in the AN languages of New Guinea.
Motu	* Ø > j / $_{\text{PrWd}}$[_a	Followed by *j > l
Cristobal-Malaitan	* Ø > j / $_{\text{PrWd}}$[_a	Followed by *j > θ; lexically gradual change.
Trukic	* Ø > j / $_{\text{PrWd}}$[_a	Homorganic glides not reported before /i/ in all languages;
	* Ø > j / $_{\text{PrWd}}$[_ {i,e} * Ø > w / $_{\text{PrWd}}$[_{o,u}	glide before /a/ is not homorganic; evidence for long-term diffusion across dialect chain.

(33) Glide accretion outside of Oceanic

Buli, Numfor	* Ø > j / $_{\text{PrWd}}$[_a	Blust (1978)
Bonfia	* Ø > j / $_{\text{PrWd}}$[_a	And other languages of Central Moluccas (Stresemann 1927: 114ff.)
Sepa, Tehoru	* Ø > j / $_{\text{PrWd}}$[_a,	Collins (1982); limited to nouns, but verbs typically take proclitic subject markers, eliminating /a/-initial verb stems (Blust 1990: 17).

Blust (1990) notes that similar sound changes appear to have occurred independently in other Austronesian languages, including Buli and Numfor, two South Halmahera–West New Guinea languages (SHWNG), and in three Central Malayo-Polynesian languages of Seram in the Central Moluccas: Bonfia (a.k.a. Masiwang), Sepa, and Tehoru. These proposed sound changes are summarized in (33).

What is of particular interest in the data summarized in (32) and (33) is the asymmetry between reflexes of **a*-initial words, which show epenthetic initial glides, and reflexes of **u*- and **i*-initial words, which, with the exception of some of the Trukic languages, do not show epenthetic glides. In other words, in word-initial position, a range of Austronesian languages show the general pattern in (34).

(34) General pattern of C-insertion under *j*-accretion
- a. ${}_{\text{PrWd}}$[a > ${}_{\text{PrWd}}$[ja
- b. ${}_{\text{PrWd}}$[i > ${}_{\text{PrWd}}$[i
- c. ${}_{\text{PrWd}}$[u > ${}_{\text{PrWd}}$[u
- d. ${}_{\text{PrWd}}$[e > ${}_{\text{PrWd}}$[e, Ø
- e. ${}_{\text{PrWd}}$[o > ${}_{\text{PrWd}}$[o (Oceanic only)

If the function of [j] in these cases is to provide an onset for the syllable in question, then why is the sound change limited to the context in (34a)? And why is the segment [j] instead of [w] or a laryngeal? I suggest that glide accretion has both a phonetic and an analogical component, as sketched in (35). On the phonetic side, glide accretion can be viewed as the phonologization of a very common sound sequence in Oceanic, /...i#a.../, realized phonetically as [ija].[15] A range of prenominal particles end in /i/. These include: generic locatives **i*, **di*, directional **ki*; personal article *i/**si*; genitive marker **qi*/**ni*; deictics **qani*, **ini*, **idi*. When pronounced as proclitics, these particles will give rise to glide-like percepts before following /a/-initial nouns.[16] The excrescent glide is analyzed as part of the following word, with *a*/*ja* as non-contrastive variants. The non-contrastiveness of these variants is due to an inherited property from Proto-Austronesian, where /j/ was contrastive word-medially and finally, but not word-initially. The system stabilizes as *ja*-forms gradually replace *a*-forms.

(35) Hypothesis regarding the origins of *j*-accretion
- a. Common sound sequence: /...i#a.../ realized as [...ija....]
- b. [...ija....] analyzed as [...i#ja...]
- c. [...ija....] analyzed as /...i#ja.../; [a....] analyzed as /ja.../
- d. /ja.../ variant chosen as basic for words beginning in [a...]

As with other examples of C-epenthesis, no direct reference is necessary to the onset-filling function of these developments. In fact, as illustrated in (34), the majority of V-initial words are inherited without change. And no direct reference is made to segmental markedness: the fact that the epenthetic glide is [j] is attributed to the high frequency of phonetic *ija* in sandhi contexts.

(ii) What is Ritwan *l*-sandhi?

Sound changes similar to *j*-accretion occur where phonetic glides are just one variant in sandhi contexts. Recall the evidence from Ritwan languages Wiyot and Yurok for original V-initial words, with *j*-insertion in sandhi (19b), and [h] inserted at the beginning of a prosodic word. In this case, an additional sandhi process involves surface

[15] Final C-loss in many Oceanic languages (29) will make hiatus across word boundaries much more common than in other Austronesian languages which maintain final consonants.

[16] Notice that in some languages, like Sepa and Tehoru, *j*-accretion is limited to nouns. The proposal sketched in (35) is not meant to account for homorganic glides in Trukic, which have a natural history similar to that outlined in 4.2.1. It will also not account for the * Ø > *w*/${}_{\text{PrWd}}$[_V in Chamorro, which occurred at some stage after **h* loss. Since Chamorro did undergo the unconditioned sound change of **e* > *u*, it is tempting to relate *w*-epenthesis to the greater frequency of final *u*'s in hiatus contexts.

[l]. Following Blevins and Garrett (2007), certain *l*-final particles, including the locative **tol*, were subject to *l*-devoicing and loss before consonants, with [l] surfacing only before vowel-initial words. This *l*-sandhi has given rise to distinct sound patterns in Yurok and Wiyot.

The Wiyot pattern is illustrated in (36): in Wiyot, a stem with an initial /h/ appears with [j] instead after preverbs ending in front vowels, and with [l] after all other preverbs (Teeter 1964: 24; Reichard 1925: 19). These sources are abbreviated T and R respectively below, with page numbers following.[17]

(36) Wiyot *h*-sandhi

a. h → j / {i, e} # ____

hakʷt-	'build fire'	bas hi jákʷtad	'then one builds a fire'	(T 114)
haʔlab-	'dance'	ki jaʔlabìɬ	'they never danced'	(T 119)
hil-	'say'	hi jíliɬ	'then he says'	(T 109)

b. h → l / {a, o, u} # ____

hap-	'be cooked'	kitko kowa láp	'they are almost cooked'	(T 117)
haʔlab-	'dance'	pitabaláʔ labiɬ	'he only dances'	(T 119)
halok-	'go along'	to lalókiɬ	'he goes along'	(T 120)

c. h → l / C # ____

hoɬb-	'feel so'	kuc kóbaɬ loɬbiɬ	'he felt bad about that'	(T 111)
hanelis-	'arrange'	ku-cap-la:nelis-oiʔ	'were arranged same way again'	(R 63)
hil-	'say'	kwis-le:l-iɬ	'suddenly he said'	(R 52)

In Yurok until the mid-twentieth century, /h/-initial words were pronounced with initial [l] only after particles which contained historical final laterals; elsewhere we find [j] after /i/, [ɣ] between other vowels, and [h] elsewhere. Examples of the general Yurok pattern are given in (37).

(37) Yurok general *h*-sandhi (Robins 1958: 9)

a. h → j / i #____

heɣo'l	'he goes'	ni jeɣo'l	'he goes there'
hunowoni	'growing'	k'i junowoni	'things that grow'

b. h → ɣ/ V #____ (V≠i)

hohkumek'	'I work'	me ɣohkumek'	'I worked'
hoole'meɬ	'they go'	wonu ɣoole'meɬ	'they went up'

Yurok examples of postparticle [l] are given in (38).

(38) Yurok *h*-sandhi: [l] after *ʔo* 'locative', *ma/me, ʔema/ ʔeme* 'past', *tem(a)* 'in vain'

FUT = future, INT = intensive, LOC = locative, PAST = past, TEMP = temporal

a. Sample place names

haʔaaɣ	'rock'	ʔo laʔaaɣ	'LOC rock'
heɣ-	'go, travel'	ʔo leɣ	'where one goes'
ho'monoʔ	'tan.oak'	ʔo lo'monoʔ	'LOC tan.oak'

[17] Since Wiyot *i* and *e* both correspond to Yurok *i* (Yurok *e* = Wiyot *a*), *h* → *j* sandhi can be said to be regular before the reflexes of Ritwan non-low front vowels.

b. tuʔ witu meɬ mi woʔ ʔo leɣohku niiɣem
and for that reason not they LOC make.INT obsidian
'That is why they do not make obsidians there.' ([ALK 75.8]; tr. *YM* 436)

c. ʔo le'm kwilek nek ki nepaane'm ko ʔo lewoloce'm
LOC say.3SG well me FUT eat.2SG TEMP LOC get.well.2SG
'It said, "If you eat me, you will recover." ' ([ALK 75.25–26]; tr. *YM* 313)

d. tuʔ hii, toʔ kwilek me lego'l mewimor
and *hii* and well PAST go.3SG old man
'*Hī*, the old man is the one who was there then.' ([ALK 75.8]; tr. *YM* 436)

A summary of the analysis detailed in Blevins and Garrett (2007) is given in (39).

(39) Ritwan sandhi before *V

Stage I Merger of laterals **l* and **ɬ* in final position led to phrasal alternations

a. Before a vowel-initial syntactic word in the same phonological word: *-V**l**#V-, i.e., [-V.**l**V-] with syllabification into the following onset

b. Elsewhere: *-Vɬ. (e.g., [Vɬ.CV])

Stage II The automatic *h*- in vowel-initial words transformed these phrasal alternations as follows:

a. Before a vowel-initial word: *-V.lV-

b. Before a consonant-initial word: *-Vɬ.CV-

c. An *h*-initial word with no lateral sandhi: *hV- phrase-initially, *ɣV- medially

Stage III Final *l* was reinterpreted as part of a following, originally vowel-initial word.

Yurok: Within the phonological word, initial *h* → *l* after certain words.

Wiyot: Within the phonological word, initial *h*, non-initial *l*.

The purpose of this somewhat long excursus on Ritwan sound change is to highlight two points. First, reanalyses where word boundaries are aligned with syllable boundaries are not uncommon, and seem particularly likely where the particle + noun combination is of very high frequency. Second, where such restructuring occurs, the segment in question may be a result of phonetic conditioning (Oceanic *j*-accretion), phonologically conditioned allomorphy (e.g., English *a* / *an*), or a combination of phonetic, phonological, and morphological conditioning, as in the Ritwan developments just discussed.

In the analyses of Oceanic *j*-accretion and Ritwan *l*-sandhi, there is no reference to segmental or syllabic markedness constraints. As (34) illustrates, *j*-accretion did not fulfill this general function, nor is there evidence for the majority of Oceanic languages that any other consonant did. In Wiyot, prior to general *l*-sandhi, all syllables already had onsets. The shift of **h* > *l* under sandhi appears to constitute a change from a less marked to more marked consonant in this environment. However, this is precisely the pattern of other cases of rule inversion seen in 4.3.2 where consonant∼zero alternations may involve segments other than glides and laryngeals. Under the current analysis, shifts like **h* > *l* in Wiyot are expected where high-frequency surface sound patterns are involved.

4.4 OTHER PLACES WHERE SEGMENTAL AND SYLLABIC MARKEDNESS FAIL

The markedness constraints in (2) and (3) have been claimed to play a role not only in consonant epenthesis, but also in sound change, affix positioning, and reduplication. In this section I briefly summarize evidence against these constraints in each area.

4.4.1 General syllabification and initial C-loss

If markedness constraints like those suggested in (2) and (3) play an active role in sound change, as argued, for example in Kiparsky (1988, 1995, this volume), then we should see evidence of this in the historical record. I have argued above that once synchronic consonant epenthesis alternations are deconstructed, the independent components which give rise to them can be stated without reference to segmental or syllabic markedness. Another argument against the role of ONSET as a constraint on sound change, however, is the existence of recurrent sound changes leading to onsetless syllables, and their apparent consequences.

In Blevins (2001) I summarize data on initial consonant loss in dozens of Australian languages, and argue for independent parallel developments in at least four distinct subgroups. The general sound change, stated in (40), involves loss of an initial consonant. While in some cases, the quality of the consonant arguably plays a role, in Arandic and Northern Paman languages, loss is prosodically conditioned with all initial consonants succumbing.

(40) Initial C-loss in Northern Paman and Arandic
(Hale 1962, 1964; Koch 1997; Blevins 2001)
* C > Ø / $_{\text{PrWd}}[_$

Since this sound change in its numerous instantiations appears to be eliminating precisely the preferred syllable types which constraints like (2) attempt to enforce, something more clearly needs to be said. But once we admit that consonants can be phonetically weak in initial position, and perhaps not accurately perceived under destressing, while at the same time admitting that hiatus contexts can result in the percept of an intervening glide in cases of historical glide epenthesis, what arguments remain for ONSET as a component of universal grammar?

One might argue, following Jakobson (1929/1962) and Kiparsky (1995, this volume), that there are simply certain universals which are never violated. Onset in (2), however, is not one of them. One of the most interesting aspects of the Northern Paman and Arandic languages which have undergone the sound change in (40) (as well as final vowel loss) is the extent to which their syllabification algorithms are highly aberrant from a cross-linguistic perspective. Sommer (1969, 1970) argues that Oykangand, which has only vowel-initial words, syllabifies all medial consonants in the coda, as schematized in (41). A similar argument is made by Breen and Pensalfini (1999) for Arrernte.

(41) Syllabification in Oykangand and Eastern Arrernte

VCV	→	VC.V
VCCV	→	VCC.V
VCCCV	→	VCCC.V

In word-based syllabification models like that advocated by Steriade (1999a) and Blevins (2003a, 2003c), the syllabifications in (41) are precisely those expected when word forms happen to converge on being vowel-initial and consonant-final.

4.4.2 The non-emergence of the unmarked I: infixation[18]

Theory-internal arguments have been made within Optimality Theory that both the position of infixation and aspects of reduplicative phonology follow from syllabic and segmental markedness constraints, which may be invisible in other processes due to overriding faithfulness constraints (Prince and Smolensky 1993; Kager 1999; McCarthy 2002). Predicting the position of infixation seems a non-issue. As argued by Blevins (1999) for Leti, and more generally by Yu (2003), there are cases of infixation which defy syllable markedness accounts, as well as languages with minimal prefix/infix or suffix/infix pairs. Data from Leti illustrating the inutility of an ONSET constraint is shown in (42).

(42) The eight allomorphs of the Leti nominalizing (NOM) morpheme

	NOM	Verb stem	Gloss	Derived nominal	Gloss	Affix type
a.	-ni-	kaati	'to carve'	k-ni-aati	'carving'	infix
b.	-n-	kini	'to kiss'	k-n-ini	'kissing'	infix
c.	-i-	mai	'to come'	m-i-ai	'arrival'	infix
d.(i)	i-	atu	'to know'	i-atu	'knowledge'	prefix
d.(ii)	ni-	atu	'to know'	ni-atu	'knowledge'	prefix
e.	ø	ruru	'to tremble'	ruru	'trembling'	null
f.	nia-	ltieri	'to speak'	nia-ltieri	'speech'	prefix
g.	i-,-i-	natu	'to send'	i-n-i-atu	'sending'	prefix + infix

Since /n/ is a potential allomorph of the nominalizing infix, we expect it before vowel-initial stems like /atu/ since, precisely in this position, /n/ will result in an onset for the vowel-initial syllable without resulting in hiatus. However, as shown by the forms in (42di), (42dii), the two attested forms are inconsistent with ONSET as a driving force in infix-placement.

A minimal prefix/infix pair from Atayal (Egerod 1965; Yu 2003) is shown in (43), the reciprocal/reflexive shows an *m*- prefix, but the actor focus, an -*m*- infix. Clearly, the position of infixation cannot be accounted for in terms of prosodic constraints, unless these constraints are specific to the morphological construction involved. This

[18] The non-emergence of the unmarked (TNETU) effects are generally unremarkable within Evolutionary Phonology, since there are no markedness constraints. The points raised here are meant to highlight just some of the empirical difficulties of claimed TETU effects.

is equivalent, however, to specifying whether the morpheme is a prefix or infix, greatly weakening the general claims of the phonologically based account.

(43) Actor focus and reciprocal/reflexive /m/ in Atayal

Reciprocal/reflexive	Actor focus	Root	Gloss
mkaial	kmaial	kaial	'talk'
mqul	qmul	qul	'snatch'
msbil	smbil	sbil	'leave behind'
mspuŋ	smpuŋ	spuŋ	'measure'
msuliŋ	smuliŋ	suliŋ	'burn'
mhkaŋiʔ	hmkaŋiʔ	hkaŋiʔ	'search'

It is also worth mentioning in this context that a strong prediction of the syllabic markedness account of infixation is unattested. Consider a language with an affix /m-/ which is generally aligned with the beginning of the word (i.e., is a prefix). Now imagine that in this same language, the only relevant constraint is ONSET. The hypothetical pattern is illustrated in (44). If an onsetless syllable occurs anywhere in the stem, the affix fills the empty onset slot (44a–d), and if there is more than one onsetless syllable (44d), the affix occupies the onset position which is closest to the beginning of the word. The advance of the /m-/ affix within the word in (44) subject to the neediness of onsetless syllables might seem an absurd sound pattern to students of historical linguistics, since it is unclear how such a pattern could arise. However, this is precisely the pattern predicted by displacement models where the actual position of an affix can be underdetermined (left edge vs. right edge), with syllable markedness constraints, among others, leading to its correct positioning. In fact, the pattern in (44) is expected to be quite common, since the only constraint which is violated is affixal alignment, precisely the constraint which is violable in other analyses where infixes are analyzed as displaced prefixes or suffixes. To my knowledge, however, there is no natural language which instantiates the general pattern illustrated in (44). This is unsurprising within the Evolutionary model, where no universal onset constraint is posited, and where no natural sound change, sequence of sound changes, inversion of a sound change, or analogical change, is known which could produce the affix distribution in (44).

(44) Hypothetical /m-/ (Align left, with ONSET undominated)

a.	m + /alu/	**m**alu
b.	m + /talua/	talu**m**a
c.	m + /talukia/	taluki**m**a
d.	m + /taukia/	ta**m**ukia
e.	m + /talu/	**m**talu, t**m**alu or ta**m**lu (depending on ranking of *COMPLEX, NOCODA, etc.)

4.4.3 The non-emergence of the unmarked II: reduplication

The other phonology/morphology interaction where unmarked structures have been claimed to be emergent is in reduplication (McCarthy and Prince 1994, 1995).

Theory-internally, of course, it is possible to analyze nearly any sound pattern as the output of appropriately ranked markedness and faithfulness constraints within Optimality Theory. The reduplicative patterns in this section are offered as examples where, whatever technical fixes one chooses to make, unexpected or "marked" sound patterns arise *only* under reduplication. These patterns, like the sound change in (40), the emergent syllabifications in (41), and the infixation patterns in (42) and (43), arguably constitute examples of marked structures emerging precisely where Optimality Theory predicts unmarked structures.[19]

In (45) Southern Oromo (Stroomer 1987) reduplication is illustrated. The general pattern is to take the first CV of the base, followed by an epenthetic *m*.

(45) Southern Oromo reduplication in frequentative verbs: CVm- (Stroomer 1987)

Base	Reduplication	Gloss
eege	**em**eege	'he waited long'
harkifte	**ham**harkifte	'he pulled frequently'
teece	**tem**teece	'she sat down a long time'
fuugite	**fum**fuugite	'she raised some children'
dubbane	**dun**dubbanne	'we talked a long time'
bak'atani	**bam**bak'atani	'they ran and ran'
deemee	**dem**deemee	'he went and went'
tataanii	**tam**tataanii	'they stayed and stayed'
guddisani	**gum**guddisani	'they made bigger and bigger (educated)'

In terms of syllable and segmental structure, the appearance of *m* is problematic. At the level of syllable structure, closed syllables are marked in comparison with open syllables. Open syllables then should be emergent, all else being equal. And, as discussed in 4.3.2, epenthetic segments like *m* pose difficulties for segmental markedness accounts: there is no sense in which *m* is natural or predictable in this context, nor is there any sense in which it is in general a natural or predictable default coda or onset segment.[20]

In (46) Trukese verbal reduplication is illustrated. Like the Oromo examples in (45), an unexpected segment and syllable type occurs in a context where unmarked structures are predicted to emerge. In (46b) and (46d), where etymological vowel-initial stems are involved, a geminate [kk] surfaces under reduplication. As with the Oromo example, this pattern is marked in both syllabic and segmental terms.

[19] For additional problematic cases of reduplication where marked structures emerge and for alternative explanations of these patterns, see Blevins (2003d, 2005b).

[20] Since coda nasals can be potentially weakened and lost, a preliminary hypothesis is that an original *CVm-CVm.C... pattern of reduplication was realized as *CVm-CV:.C... by a prosodically conditioned change of this type. The *CVm- pattern was then extended to all CV:-initial stems, and later to other stems as well. While this suggestion is purely hypothetical, it illustrates the usefulness of Evolutionary Phonology in limiting the choice space for historical development. One cannot simply assume that in Southern Oromo, *m* is the unmarked consonant, and one cannot motivate the occurrence of /m/ by claiming that the prefix must be a closed syllable, since in the *emeege* example, it is not. On the other hand, one cannot motivate the insertion of *m* through ONSET, since this constraint plays no role in reduplication of C-initial forms.

Syllabically, initial geminates are marked in comparison with non-geminates. Segmentally, epenthesis of *k* is uncommon, and lacks phonetic motivation.

(46) Trukese reduplication (Goodenough and Sugita 1980) HAB = habitual

	Non-red. Base (N or V)	Single C- RED (derivational)	Double CV-C- RED. (productive,) (inflectional)	Gloss
a.	fótuki	ffót		plant it/be planted
	posuuw	ppos		stab him/be stabbed
	tuuna-	ttuun		twirl/be twirled
b.	áápi	kkááp		transport it/transporting
	éréétí	kkéréét		scrape, sand/be sanded
	amaat	kkamaat		be ground/grinding
c.	suku-	ssuk	sussuk	knock/knock (rep.)/HAB
	fene-	ffen	feffen	peck/peck (rep.)/HAB
	tuu-	ttu	tuttu	stab, pierce/be sewn/HAB
	pwúnúwa-		pwúppwúnú	spouse/treat as a spouse/HAB
d.	wún		wúkkún	drink/HAB (POC **inum*)
	ósómwoonu		ókkósómwoonu	pay chiefly respects to/HAB
	eesa		ekkees	son-in-law/treat as s-i-l HAB

For Oromo, we can only guess at the sequence of historical developments. But the historical phonology giving rise to the Trukese patterns in (46) is well documented. The single C-reduplication pattern reflects earlier CV- with loss of the pretonic unstressed vowel (Goodenough and Sugita 1980; Blust 1990). What complicates the sound patterns is the loss of word-initial **k*, but maintenance of word-initial **kk*. Reduplicative paradigms after **k*-loss show V-initial forms in the base, and geminate *kk* under reduplication. Of particular interest in this case is the extension of the geminate *kk* pattern to vowel-initial words which did not have an etymological initial **k*. In this class is the verb *wún* 'drink', from Proto Oceanic **inum*, as well as many others.

4.5 SUMMARY AND IMPLICATIONS

In this chapter, I have suggested that there are clear natural and unnatural histories for patterns of consonant insertion which make no reference to syllable onset or segmental markedness. At the same time, I have offered new ways of understanding the typology of C-epenthesis. Within the realm of natural history, glide epenthesis and laryngeal epenthesis are two distinct subtypes with different phonetic and phonological profiles. In the domain of unnatural histories, significant correlations are observed between consonants subject to coda weakening and those involved in epenthesis. This finding follows from our understanding of rule inversion as part of phonological

acquisition. Finally, a mix of natural and unnatural history characterizes the analysis of Oceanic *j*-accretion and Ritwan *l*-sandhi.

What are the implications of this study for phonological modeling? Within the phonological realm, there appear to be few, if any, substantive universals.[21] Universal tendencies emerge from recurrent instances of phonetically natural sound change, and from common events like rule inversion which have no phonetic basis. Natural and unnatural histories are intrinsic aspects of language change and give rise to synchronic systems in which the contributions of each are superficially indistinguishable. Synchronic consonant epenthesis is not a monolithic entity, and we are no closer to understanding it by simply listing and formalizing every case which occurs. However, by shifting our focus from synchronic universals to common and recurrent trajectories in sound change, the distinct histories which characterize superficially indistinguishable sound patterns may be disentangled, with explanations for universal tendencies embedded within them.

[21] Though, see Kiparsky (this volume) for a different view. The widely held view that distinctive features are substantive phonological universals is undermined by Mielke (2004), where features are argued to be language-specific emergent properties of grammars, defining both phonetically natural and unnatural classes.

5

Formal Universals as Emergent Phenomena: The Origins of Structure Preservation

Joan L. Bybee
University of New Mexico

5.1 INTRODUCTION

All explanations for linguistic phenomena, both universal and language-specific, must necessarily have a diachronic dimension, since all linguistic phenomena have histories which determine their present conventionalized state. With respect to language universals—more appropriately called "cross-linguistic similarities" since there are so few absolute universals—I have argued that an explanation is not valid unless it can be demonstrated that the explanatory principle is actually at work in the mechanism of change that brings about the cross-linguistic pattern (Bybee 1988b). Taking the role of diachrony one step further, one could argue that since there are so few absolute universals, identifying the mechanisms of change behind cross-linguistic patterns will lead us closer to an understanding of the factors that produce cross-linguistic patterns, and these factors, I would maintain, are the only true universals of language in the sense that they operate in all languages at all times.

Thus, the focus for establishing the explanations for cross-linguistic similarities should be on the mechanisms of change (Bybee et al. 1994; Bybee 2006b). Identifying the causal mechanisms in change requires a detailed look at all the properties of a change—including its directionality, gradualness, spread through the community and through the lexicon—as these properties may give clues to the mechanism involved. For instance, the lexical diffusion of a phonological change gives important clues as to its causes: change taking place in phonetic environments affects high-frequency words before low-frequency, pointing to the automatization (with practice) of the neuromotor sequences involved in the production of the word; in contrast, change that leads to the regularization of paradigms affects lower-frequency items first, pointing to

the mechanism of analogical change used when low-frequency forms are not readily accessible (Hooper 1976b; Phillips 1984; Bybee 2001).

A central tenet of Usage-Based Theory is that structure is created as language is used. In the preceding example, both neuromotor practice and analogy are processes that occur in individual usage events. With multiple applications of a mechanism within and across individuals, a change might progress to the extent that it is noticed by linguists and by speakers.

What are the stages in a "usage event" where change might occur? These include the selection of expressions, lexical access by the speaker, articulatory production, perceptual decoding, lexical access by the hearer, categorization, assignment of meaning and inference-making. All of these operations have certain inherent tendencies towards change, especially upon repetition. With repetition by different speakers, these tendencies can develop into a noticeable linguistic change (Bybee et al. 1994; Bybee 2001; Pierrehumbert 2001; for the speaker as the locus of language change, see Keller 1990/1994 and Croft 2000).

As change is initiated and carried out by the same mechanisms across languages, we find very similar paths of change in related and unrelated languages at different times. This has been noted in phonology (Foley 1972; Mowrey and Pagliuca 1995; Blevins 2004a), where, for instance, a common path of change for voiceless labial [p] is cross-linguistically documented as [p] > [pf] > [f] > [h] > zero. In the grammatical domain, such paths of change are also amply documented: e.g., constructions signaling movement towards a goal become futures, verbs meaning 'finish' become perfects and pasts, a coordinate clause can become subordinate, a verb becomes an auxiliary, and so on (Lehmann 1982/1995; Heine and Reh 1984; Heine et al. 1991; Hopper and Traugott 1993; Bybee et al. 1994).

The Greenbergian theory of language universals (Greenberg 1969, 1978a, 1978b) views language as a complex system. The synchronic cross-linguistic patterns are not the end point of universals research, but just the starting point: synchronic patterns are the result of movement along these common paths and underlying the paths are certain recurring mechanisms of change, which have the following properties:

1. mechanisms of change are universal in the sense that they can be found operating in all languages at all times;
2. they are relatively few in number;
3. they involve neurocognitive tendencies that manifest themselves as language is produced and processed;
4. they apply during individual usage events; and
5. the cumulative effect of their application over multiple usage events creates grammar.

This view is consonant with the theory of complex systems, in which the systematic structure of language is considered to be continually evolving through the ongoing application of processes during multiple usage events. Grammar (the cognitive

organization of language) is thus said to be "emergent" rather than fixed. The ability to create language systems through categorization, analogy, neuromotor automatization, semantic generalization, and pragmatic inferencing derives from the innate neurocognitive capacities of human beings. These are largely domain-general capacities that happen to be used to create language. The hypothesis is that there is no need to posit innate linguistic universals, but rather that the similarities that exist across languages can be explained through the interaction of a small number of mechanisms of change.

The complex system view contrasts with that of Kiparksy (this volume), who distinguishes between patterns created by change and generalizations written into Universal Grammar (or innate generalizations). Note that Kiparsky's theory considers some cross-linguistic generalizations to be universals of grammar, while the argument to be pursued here is that the deeper level of explanation requires understanding the mechanisms of change.

The explanatory power of diachronic typology is also demonstrated in the chapters in this volume by Hopper and by Kuteva and Heine. Hopper demonstrates that an understanding of a well-established grammatical pattern in many languages (verb serialization) can be fruitfully studied in languages where it is only a minor tendency (such as English) and that a thorough, discourse-based analysis sheds light on the origins of the construction type. Kuteva and Heine show that given the set of mechanisms behind grammaticization, both generalizations and exceptions can be explained. While the mechanisms are applicable in all languages at all times producing the common paths of change as illustrated above, these mechanisms also sometimes produce other outcomes, making it possible to have other, minor paths of change as well, depending upon their interaction and the type of linguistic material they apply to.

In this chapter, I will illustrate the relationship among synchronic universals, paths of change, and mechanisms of change with respect to the phonological changes that create the structural tendency known as Structure Preservation in Lexical Phonology. The outline of this explanation is given in Bybee (2001: 214–215), but here it is worked out in more detail.

5.2 SUBSTANTIVE AND FORMAL UNIVERSALS

Substantive universals are those cross-linguistic tendencies that involve either phonetic or semantic substance, while formal universals are those tendencies that involve grammatical form or the structure of the grammar (Chomsky and Halle 1968). Paths of change can be categorized in the same way. Paths that specify changes in phonetic substance, such as the reduction of a voiceless labial stop shown above, are substantive. Within the framework of grammaticization, examples of substantive universals are those paths of change involving meaning, such as the generalization that anteriors

(perfects) become perfective or past. A parallel formal universal involves the cline discussed in Givón (1979) and Hopper and Traugott (1993), by which a content item becomes a grammatical word, then a clitic, and then an affix. When these substantive and formal paths of grammaticization operate simultaneously, the result is a perfective marker that is an affix. The operation of these two paths accounts for the fact that with few exceptions, the perfective and past are marked with affixes (Bybee and Dahl 1989; Bybee et al. 1994).

However, formal universals could also refer to the form of the grammar, as in properties such as modularity. Such properties are usually considered to be given innately as a starting point for language acquisition (Kiparsky, this volume). However, it is also possible that the general structure of modularity is emergent from the nature of change. That is, certain recurring, parallel paths of change create patterns that are largely modular. Under this view, there would be transitional phases between modules, i.e., exceptions to the strict separation of levels. Given that there is ample evidence that such exceptions exist—phonological alternations dependent upon morphology and syntax, as well as morphological and syntactic alternations dependent upon phonology—examining an emergentist view of such separations becomes a necessary endeavor.

The main focus of the current chapter is the principle of Structure Preservation, which deals with the distinction between contrastive and non-contrastive segments and has been formulated as a structural universal of language (Kiparksy 1985). By examining a case which creates difficulties for this principle, I show that this proposed structural universal is in fact emergent from the parallel development of three unidirectional paths of change, propelled by certain mechanisms of change, which are universals in the sense that they apply in all languages at all times.

5.3 A FORMAL UNIVERSAL: STRUCTURE PRESERVATION

Structure Preservation is a principle formulated in Lexical Phonology (Kiparsky 1985), though it reflects a principle recognized in earlier structuralist theories. The principle states that only contrastive sounds or features take part in morphologically or lexically conditioned alternations; or, stated differently, alternations that are restricted to the word level involve only contrastive features. Segments or feature combinations that are non-contrastive must be introduced by postlexical rules, which are automatic and phonetically conditioned and often apply across word boundaries.

This principle correctly captures a strong tendency in the languages of the world for alternations conditioned either lexically or morphologically (or both) to involve contrastive features and segments. Consider for example two alternations that English /k/ enters into: in some words with Latinate affixes /k/ alternates with /s/, as in

electri[k], *electri[s]ity*; *criti[k]*, *criti[s]ism*. This alternation applies at the word level: it does not occur when two words come together; it is unproductive, lexically restricted and at least partially morphologically conditioned. In contrast, English /k/ also has a palatal variant [c] before a front vowel, as in *key*, *kiss*, *came*. This variant appears automatically (that is, the process that creates it is productive), and it is not lexically or morphologically restricted. In principle it could apply when two words come together, as in *break even*, though I know of no phonetic studies that show that this is the case. This sort of situation—where phonemes alternate when there are lexical or morphological restrictions and non-contrastive elements alternate in purely phonetic environments—is typical of the phonologies of the languages of the world. Kiparksy designates it as "Structure Preservation" because the lexical phonological rules do not introduce any feature combinations that are not already present in the lexicon. That is, a lexical phonological rule could not introduce a palatal stop into the English lexicon.

A principle with the same effect was discussed in American structuralism under the rubric of "separation of levels". The phonemes of a language together with their allophones could be arrived at using only phonetic information (Hockett 1942). Once the phonemes were established by phonetic principles such as complementary distribution, then alternations among phonemes in words could be discovered.

Early in this discussion Pike (1947, 1952) correctly noted that using only phonetic information to predict the occurrence of allophones was impossible both in terms of procedure and in terms of theory. He noted in particular that the behavior of phonemes at junctures (boundaries) could only be predicted on the basis of grammatical and lexical information, not purely phonetic information. He brings up the contrast between *nitrate* [najt^{h}rejt] and *night rate* [najt˺rejt], in which the allophone of /t/ that is used depends upon knowing that the /t/ in the latter phrase occurs at the end of a word. This case is not particularly a problem for Kiparsky's formulation as long as word boundaries can block the postlexical rule of aspiration.

Counter-examples to Structure Preservation have also been discussed (see below). We are presented, then, with a typical dilemma in linguistic theory: a strong tendency is evident in the grammars of all languages encountered; it seems to represent a basic organizing principle of language and yet, if it is canonized as a structural principle, counter-examples or exceptions quickly come to light. In the face of exceptions, researchers try to revise the principle or reanalyze the counter-examples. A question that rarely arises, however, is why grammars would have such an organizing principle. I suggest that if we take explanation as the primary goal and set about trying to understand why this strong tendency exists, and how it arises in languages, we can explain not only the general tendency but the exceptions as well, and gain further insight into the nature of grammar.

A famous counter-example to the Structure Preservation principle (discussed extensively in the structuralist and the generativist literature) is the alternation between German [x] and [ç] (Moulton 1947; Leopold 1948; Hall 1989). The basic facts are these

(examples from [Hall 1989]): [x] occurs after back vowels; [ç] occurs after front vowels and /n/, /r/, and /l/.

(1)	Buch	[bux]	'book'	siech	[zi:ç]	'sickly'
	Koch	[kɔx]	'cook'	Pech	[pɛç]	'bad luck'
	nach	[nax]	'after'	Köchin	[kœçɪn]	'cook (fem)'

However, the diminutive suffix *-chen* is always [çən]:

(2)	Kuhchen	[ku:çən]	'little cow' (*Kuh* + *chen*)
	Tauchen	[taoçən]	'little rope' (*Tau* + *chen*)
	Pfauchen	[pfaoçən]	'little peacock' (*Pfau* + *chen*)

The invariant form of the diminutive despite the preceding vowel produces phonemic contrasts with the following words:

(3)	Kuchen	[ku:xən]	'cake'
	tauchen	[taoxən]	'to dive'
	pfauchen	[pfaoxən]	'to hiss'

In addition, assimilated borrowings use the palatal fricative in word-initial position.

(4)	Chirurg	[çirʊrk]	'surgeon'
	Chemie	[çemi:]	'chemistry'
	Cholesterin	[çolɛsteri:n]	'cholesterol'
	Fotochemie	[fo:to:çemi:]	'photochemistry'

The dilemma is that one would like to analyze the velar and palatal fricatives as allophones of the same phoneme, with the palatal being the mere output of a postlexical rule (if you choose the velar as the underlying phoneme), but that pesky diminutive suffix makes such an analysis impossible. Moreover, borrowed words with the palatal fricative in word- and syllable-initial position create additional problems. In pregenerative structuralism, Moulton (1947) argued in favor of a segmental juncture phoneme preceding the /x/ to condition the palatalization. However, in the diminutive suffix, since no pause is present, this juncture has a zero realization. Leopold (1948) argues that such an analysis is circular since we only know that the juncture is there because the [ç] appears. Similar discussions in Lexical Phonology (Hall 1989 and MacFarland and Pierrehumbert 1991) also lead to lack of consensus.

This case lends itself nicely to a diachronic explanation: the alternation started out as phonetically conditioned, as the older form of the suffix *-ichiin* had the front vowel conditioning context within the suffix. Apparently the palatal variant was established before the first vowel was lost, so that it remained despite the loss of its conditioning environment. Now [ç] has gradually become established as an independent element (or phoneme) in the diminutive suffix and has also been recruited for use in loanwords. Moreover, in some dialects the palatal variant is now prepalatal [ʃ], indicating that the phonetic distance between the two originally predictable variants has also increased. The "exception" is really a kind of intermediate case, and as such has a diachronic

explanation, but does this contribute to our understanding of the synchronic principle? I will argue in the remainder of the paper that indeed it does. I will argue that the general principle is not a synchronic organizing principle of grammar, rather a general tendency that results from the coevolution of phonological changes along several parallel paths. I argue, then, that both the general principle and the exceptions to it have diachronic explanations.

5.4 THREE UNIDIRECTIONAL PATHS OF CHANGE IN PHONOLOGY

Three well-documented universal paths of change occur in parallel and lead to the synchronic situation that is described as Structure Preservation. First, phonetically conditioned sound change creates alternations that gradually acquire morphological or lexical conditioning (Vennemann 1972; Hooper 1976a; Dressler 1977, 1985).

(5) phonetic conditioning > morphological or lexical conditioning

Second, what starts as a small phonetic change tends to continue to change phonetically over time, leading to a greater distance between the original sound and the resulting one. Thus the two alternating sounds grow more different from one another (Hooper 1976a; Janda 1999).

(6) small phonetic change > larger phonetic change

For instance, a [k] before a high front vowel might move forward to a palatal position. The extent of palatalization might increase until the sound in that context becomes an alveo-palatal affricate. Such changes are documented in Romance languages, where Latin /k/, going through stages such as [tʃ], [ts], ends up as [s]. This created the alternation discussed above, between /k/ and /s/, that was borrowed into English along with the French words.

Third, simultaneous with the preceding developments, productive phonetically conditioned alternations between two sounds are likely to become unproductive. This path of change is related to the previous two in ways that will be discussed below. An example would be the loss of intervocalic voicing of fricatives in English; this process created the *wife*, *wives* alternation, but now is no longer productive, as voiceless intervocalic fricatives are allowed in English (e.g., *classes*).

(7) productive processes > unproductive

These paths of change together result in Structure Preservation, since as a change begins to take on morphological and lexical conditioning, the new variant tends to grow more distant from its source, producing a larger phonetic change, one that could be phonemic. Simultaneously, its tendency to become lexicalized and to cover a larger phonetic distance leads to the loss of productivity and the ability of the new sound to occur in contrast with the original sound. The actual mechanisms behind these paths of change are discussed in the following sections.

5.5 A MODEL OF SOUND CHANGE

This section proposes a series of steps by which sound change takes place and presents a model that accounts for the phonetic gradualness of sound change, the lexical gradualness of sound change, and the eventual result that only contrastive elements occur in morphologically and lexically conditioned alternations (Bybee 2001). The model is usage-based, in the sense that cognitive representations are affected by usage events and are emergent from them. In this model, a principle such as Structure Preservation is not in itself an organizational principle of language, but rather the result of the interaction of the more basic mechanisms of change that are operative when language is used.[1]

In keeping with the usage-based viewpoint, sound change is viewed as the result of the reduction or retiming of gestures that occurs with automation of production in language use (Browman and Goldstein 1992; Mowrey and Pagliuca 1995). Sound change is manifested early on as variation in casual speech. Such variation is influenced by the phonetic context and eventually results in allophonic variation which can become quite stable. In this view, sound change is largely, if not wholly, phonetically conditioned and reductive (see the authors mentioned above and Bybee 2001 for more discussion). Note that it is phonetically conditioned sound change that creates allophones of phonemes; it follows then that in general the distribution of allophones can be stated in purely phonetic terms.

Since sound change occurs as language is used, sound change takes place in actual production units, i.e., words and phrases. The evidence for this claim is the fact that articulatorily motivated sound change takes place earlier in high-frequency words than in low-frequency words (Fidelholtz 1975; Hooper 1976b; Phillips 1984, 2001; Bybee 2000b, 2002).[2] In order to account for this lexical diffusion phenomenon, the immediate effects of sound change are registered in an exemplar representation (Johnson 1997; Pierrehumbert 2001, 2002). Exemplar representations allow a cluster of phonetic variants for a word and this cluster is constantly being updated as new variants are experienced. Cole and Hualde (1998), and Booij (to appear) argue that the fact that the effects of sound change are never reversed provides evidence for the hypothesis that sound change has an immediate and permanent effect on the memory representation of words. These researchers point out that when a sound change or the alternation it sets up has ceased to be productive, the change is not undone, as one might expect in a theory in which underlying forms remained unchanged and only surface forms are affected by the sound change (i.e., where sound change is rule addition). Thus, Booij points out that long vowels created by lengthening in open syllables in

[1] Blevins (2004: 244ff.) also notes that Structure Preservation can be derived from phonologization. My account here differs from hers both in fleshing out the details and also in the mechanisms of change that are proposed.

[2] As Hooper (1976b), Bybee (2001), and Phillips (1984, 2001) have shown, changes that affect low-frequency words first are not due to articulatory reduction, but result from other mechanisms of change.

Dutch do not shorten again when this rule becomes unproductive; rather, they stay long.

Other evidence that the results of productive processes are registered lexically is that the phonetic form of existing words can be used in the creation of new words, a phenomenon which would not be possible if only phonemic forms were stored in memory. Steriade (2000) (also arguing against a strict distinction between phonetic and phonological features) notes the difference in the medial coronal consonants in *fatalistic*, which has a flap, and *positivistic*, which has a [t]. The difference corresponds to the pronunciation of the base word, *fatal* with a flap, and *positive* with a [t]. This distinction suggests that the mental representation of *fatal* has a flap in it, as does the experimental evidence of Connine (2004).[3] Similarly, in compounds such as *night rate* the [t] is phonetically the same as it would be if it were word-final. Of course, one can derive this effect by placing a word boundary in the compound, but what it really means is that the compound is formed by using the phonetic shape of the word *night* rather than some more abstract phonemic shape.

Immediate registration of sound change in words also accounts for the tendency for phonetic change to become lexically and morphologically conditioned, as we will see in section 5.7.

If a change is occurring in a number of words, the general neuromotor routine that governs the gestural sequence is gradually changing, too. This accounts for the automatic nature of phonetically conditioned alternations, that is, the fact that they apply to new or nonce words. The general neuromotor routine itself is not static, but allows for a range of variation and may be biased towards lenition or anticipation of gestures. (See Pierrehumbert 2002.)

5.6 CATEGORIZATION OF PHONETIC VARIANTS

According to Miller (1994), phonetic variants are categorized by phonetic similarity and organized around a best exemplar, or the variant that speakers judge to best represent the category. Speakers can make such judgements appropriate to different phonetic contexts (e.g., English [t] after [s] vs. English aspirated [t^h]), suggesting that phonetic categories may correspond more to "allophones" than to "phonemes". Over multiple instances of exemplar categorization, a continuous parameter with a bimodal distribution can sharpen and separate into distinct categories. Wedel (2006) points out that in an exemplar model that includes both perception and production and models sound–meaning correspondences, overlapping or intermediate stimuli tend to be lost because their categorization is less consistent. Stimuli or tokens close to the centers of

[3] Steriade accounts for this phenomenon by using a constraint labeled Paradigm Uniformity. My account needs no such constraint; registering the variants in memory storage has the desired effect. See Garrett (this volume) for a different critique of such proposed Paradigm Uniformity constraints.

categories are more consistently classified than tokens near the boundaries between categories; as a result two nearby categories tend to diverge from one another. In addition, since marginal and infrequent members of categories tend to be lost over time, categories that once had overlapping members can evolve into distinct categories with no overlaps.

Given these categorization effects, the range of phonetic categories used in a language is dynamic and changeable, but not infinite, giving rise to a set of phonetic categories for allophones, many of which are also linked to specific phonetic environments, and thus are considered allophonic in phonological theory. Related advantages are the resulting limited set of general neuromotor routines that are used repeatedly in different words. As Lindblom (1992) and Studdert-Kennedy (1987, 1988) point out, if each word had its own unique set of gestural features there would be a strict limit on the number of words a language could have. In order to acquire and maintain an unlimited lexicon, a constrained set of gestural configurations must be reused in the words of a language. Presumably the same would hold for the perceptual configurations. Both production and perception are made more efficient by the use of a constrained set of units for all the words of a language.

For our purposes, what is most interesting here is that new categories can be formed if phonetic variants in different contexts start to differentiate. In the formation of new contextual categories, intermediate variants tend to be lost. For example, though the American English alveolar flap was originally a variant of the /t/ or /d/ categories, it has now formed a distinct phonetic category that is contextually restricted. In these contexts the new best exemplar is the flap and a full [t] or [d] does not occur in natural speech. The current range of variation for this category contains the flap and further weakened versions of it, but excludes full [t] and [d]. One can of course produce an aspirated [t^h] in a word such as *butter*, but that is done by accessing a different category.

5.7 SOUND CHANGE HAPPENS TO WORDS

The lexical storage unit that is relevant for the phonetic categories of the language is the word or phrase. This is also the unit of production to which neuromotor routines apply. Thus words tend to have constrained ranges of variation unless they are of very high frequency, in which case they may have variants specific to certain phrases.

I have argued that sound changes that take place at word boundaries show the tendency for a word to have a small range of variation: alternations created at word boundaries tend to be resolved in favor of the variant that occurs in preconsonantal position (Bybee 2000a, 2001). For example, in the reduction of Spanish syllable-final /s/ to [h], word-final /s/ at first shows variation according to the phonetic environment, with [s] occurring before vowels and [h] before consonants. Later stages show [h] extended to the majority of word-final tokens, even those with a following vowel.

Thus *más o menos* becomes [mahomenoh] 'more or less'.[4] The representation for the word *más* at one point had a large range of variation, occurring with final [s] and final [h] and many variants in between. Since the [h] variants were more frequent, as the following word would begin with a consonant twice as often as a vowel, the more marginal [s] variants were lost and the final [h] became established as the best exemplar of the word-final category.

Since such examples are common and variations at the level of the word according to phonetic context are not common (though they do occur in special constructions or phrases, as in French liaison [Bybee 2001]), I take such cases as evidence that there is a strong tendency to keep the phonetic variation in an individual word down to a small range. Only high-frequency words such as *don't* encompass wide ranges of variation, but in these cases, the variants are restricted to certain phrases, and it can be shown that each phrase is itself behaving like a word. Thus the *don't* in *I don't know* is in a different item of storage than the *don't* in *we don't smoke*, where *don't* is functioning as a separate word (Bybee and Scheibman 1999).

Since words are the units within which sound changes are established, words containing the same morpheme in different phonetic contexts provide the locus for alternations to develop. Using the Spanish example again, there are a few nouns that originally ended in [s], whose plurals would add *–es*. Thus the singular form of the noun *voz* [bos] 'voice' would become after the change [boh] while the plural *voces* would become [boseh] retaining the [s] in a position before a vowel. As the singular and plural are distinct words, this "variation" is not resolved in the same way as variation within a single word. Rather the two allomorphs are retained unless analogical change manages to level them.

Postulating words as the units of representation in memory and a tendency for words to have a narrow range of variation explains how word-level phonological alternations develop. It also explains how phonetically conditioned alternations become lexically and morphologically conditioned, the universal path shown in (5). As phonetic variants become established in words during sound change, particular morphemes take on different forms in different phonetic contexts, which are different morphological and lexical contexts as well. When a new phonetic category is established and intermediate variants are lost, the resulting alternation is associated with certain morphemes and/or stems as much as it is associated with certain phonetic conditions. Given further changes, such as an increase in the phonetic distance between variants, the loss of productivity of the original phonetic routine, or the loss of conditioning environment, only morphological and lexical conditioning will remain viable. (See section 5.8.)

As applied to the case of the German velar fricatives, the diminutive *-ichiin* always has the fricative after a palatal vowel; thus it was always produced as [ç] once this variant entered the language. In all the words with the diminutive suffix, the [ç]

[4] Examples from transcripts of Cuban speakers in the 1970s collected by Tracy Terrell.

was firmly established. Thus when the palatal vowel was lost, the [ç] remained. Its association with this particular suffix had been long established.

5.8 FURTHER CHANGES

Once a new phonetic category is established for certain phonetic contexts and represented lexically, the further changes mentioned above can occur. First, the phonetic change itself may continue to progress, as in the case of the German palatal fricative continuing to become more fronted. Such a change could be the result of the continuation of the articulatory trend that originally set the change in motion, or it could be related to the perceptual consequences of the new categorization and transgenerational reinterpretation of the variation (Janda 1999). In either case, the phonetic distance between the original variants will continue to grow. As mentioned above, the establishment of a new phonetic category also means the establishment of a new neuromotor routine. Thus there is a neuromotor routine for producing [ç], which begins to be possible even after back vowels.

Second, the new, independent set of variants and their associated neuromotor routine can be used in new contexts, as in loanword adaptation, where, for example, the German palatal fricative is used in word-initial position. The use in new combinations has the potential for creating more instances of contrast.

Third, the productivity of the original alternation is lost as the neuromotor routines are revised and the routine for producing [ç] or [ʃ] is no longer tied to the presence of a preceding front vowel. This leaves the door open for new instances of [x] to occur after front vowels as well. Such new instances would undoubtedly assimilate to the front vowel, but not to the extent that the older reflexes did.

If the establishment of new "phonemes" corresponds to the establishment of new categories that have the potential for contrast, then the change is covert: it actually happens long before exceptions develop. Thus phonetic categories that are considered predictable in traditional analysis may have already achieved the phonetic distance and the lexical or morphological associations to become phonemic when the occasion arises. Not only is the German palatal fricative such a case, but also vowel length in English, which is used as a perceptual clue to the voicing of final consonants, even though it is still "predictable" (Bybee 2001).

5.9 THE EXPLANATION FOR STRUCTURE PRESERVATION

In the preceding sections we have established that the convergence of several factors that naturally occur in change explains the tendency for Structure Preservation to

hold; indeed, it explains the general phonetics/phonology distinction, which must be viewed as a continuum. The way these factors interact is as follows.

First, sound change, realized as gradual phonetic change, takes place in words and permanently affects their representation. Thus variants are associated with particular words, phrases, or morphological categories. Second, marginal or infrequent variants of words are lost, giving the phonetic categories a limited range of variation. Third, phonetic change continues to progress, taking the changed variants farther away from their original source. This entails the establishment of new neuromotor routines that are not necessarily dependent upon the phonetic context. It also qualifies the new variants perceptually for phonemic contrast, should the occasion arise. This scenario, then, explains how and why "word-level" phonology develops and why such phonology usually involves segments and features that are used contrastively elsewhere. However, it also explains how and why intermediate cases develop.

As Greenberg (1969: 186) says in his description of this dynamic theory: "It is not so much that the 'exceptions' are explained historically, but that the true regularity is contained in the dynamic principles themselves."

5.10 MECHANISMS: PROCESSES THAT ARE CONSTANTLY IN OPERATION AS LANGUAGE IS USED

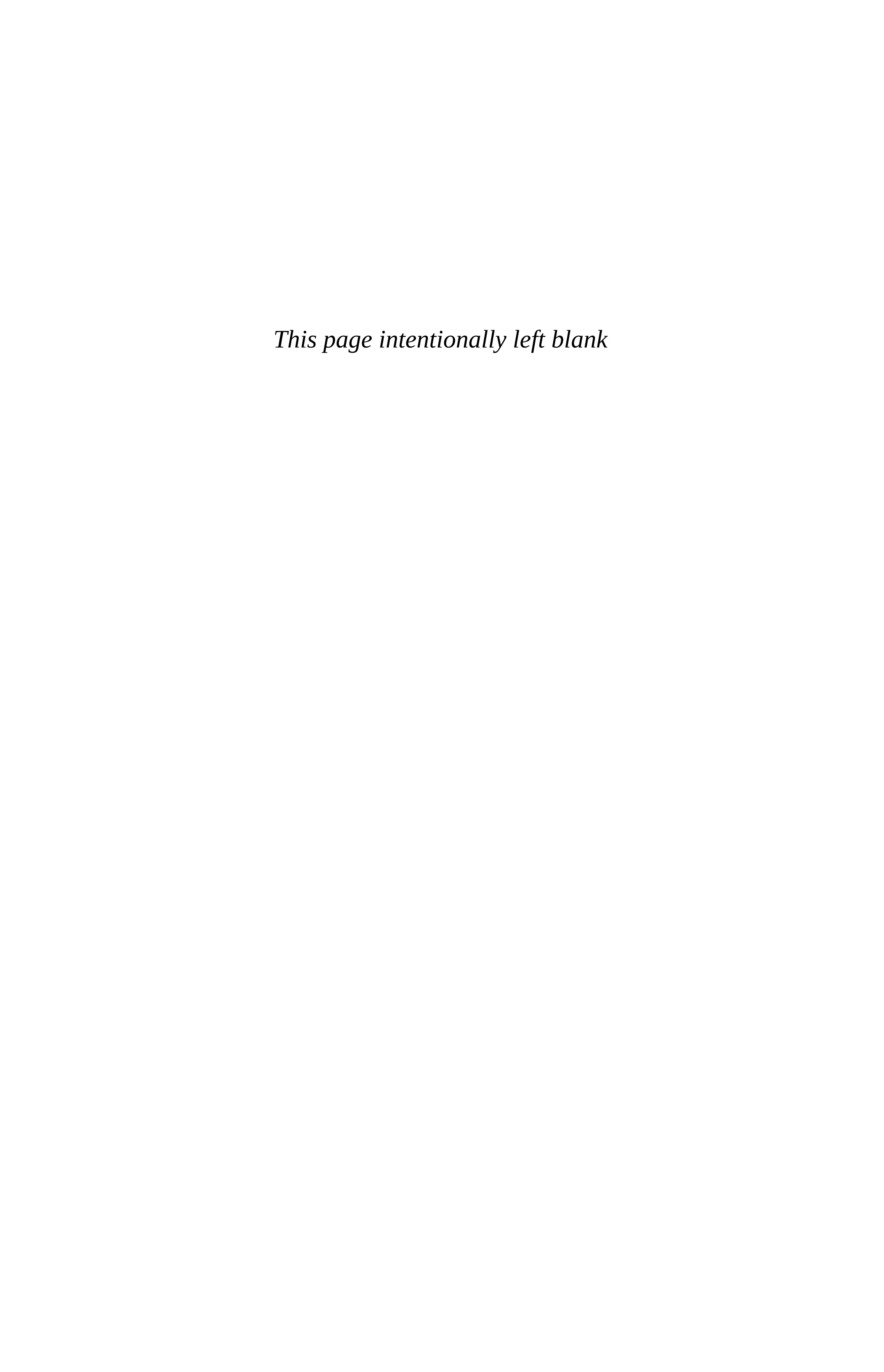
The mechanisms behind the paths of change just discussed are reviewed here.

(8) a. Repetition of sequences that make up words and phrases leads to automatization of these units and gestural reduction.
b. Cognitive representations are affected by language use; experience with language is recorded in memory; thus the effects of sound changes are registered in the phonetic representations of words immediately.
c. Phonetic variants are categorized during usage events based on phonetic similarity.
d. Repeated instances of categorization can sharpen the differences on a continuum, leading to the split of one continuum into more than one category.
e. Phonetic change in a certain direction tends to continue.

Note that none of these mechanisms that create the structure of the phonology has to be stated as a constraint. No constraints need to be formulated because the structure that evolves is a natural consequence of multiple applications of the processes that human beings use to produce and decode speech. I submit that if we begin to think realistically about the processes activated during language use, explanations for many structural phenomena will emerge.

It should be noted that structural theories, such as American structuralism and Lexical Phonology, propose no explanation for the structural properties they have

identified. A "principle" such as Structure Preservation or separation of levels is simply a property of grammars and in those theories requires no further explanation, since structure is assumed. However, a usage-based emergent grammar seeks a higher level of explanation. It is a principle of such theories that structural properties—or more appropriately, tendencies—arise as language is used and find their explanations in the nature of the categorization and processing capacities of the human brain.



This page intentionally left blank

PART III

Morphological Relationships: The Shape of Paradigms

This page intentionally left blank

6
Paradigmatic Uniformity and Markedness

Andrew Garrett
University of California, Berkeley

6.1 INTRODUCTION

Historical linguists traditionally distinguish *extension* and *leveling* as two important subtypes of analogical change. Extension is said to take place when an alternating pattern is introduced to a previously non-alternating paradigm (e.g., the irregular *drive–drove* alternation is extended in some English dialects to produce *dive–dove*); leveling is the elimination of paradigmatic alternations. A textbook case of leveling (Trask 1996: 109) is the transformation of the Old French paradigm of 'love' as in (1); the modern forms cited show leveling of the stem alternation between *aim-* and *am-*.

(1) 1SG *aim* 1PL *amons* → Modern French *aimons*
2SG *aimes* 2PL *amez* → Modern French *aimez*
3SG *aimet* 3PL *aiment*

Patterns of extension and leveling have been a testing ground for linguistic theories since the beginning of scientific linguistics (Verner 1875; Paul 1880).

In this chapter I survey a set of levelings and extensions affecting verb paradigms in two languages, English and Ancient Greek. While most of the individual changes are well described in the literature, two broader patterns have gone unnoticed. I will suggest that one pattern, shared by English and Greek, is a diachronic universal; the other sharply distinguishes Greek from English. Each pattern in turn helps answer one of the questions posed in (2) and (3).

(2) *Optimization.* Does paradigm leveling occur because of some optimizing impulse (a preference for simplicity in general or uniform paradigms in particular) or, alternatively, is it a consequence of independent mechanisms of morphological change?

For discussion and comments on early versions of this chapter I am grateful to Adam Albright, Stephen Colvin, Susanne Gahl, Larry Hyman, Sharon Inkelas, Theresa McFarland, Anna Morpurgo Davies, Pawel Nowak, Calvert Watkins, three anonymous referees, and audiences at Berkeley, Harvard, and Oxford.

(3) *Directionality.* In cases of paradigm leveling, where one alternant (e.g., stem allomorph) replaces another, what principles explain the directionality of leveling? Why is one alternant generalized while another is replaced?

These are crucial questions for any account of paradigm structure and diachrony.[1]

The optimization question in (2) bears directly on the theme of this volume. In numerous languages many paradigms are non-alternating; language change often transforms alternating paradigms into non-alternating ones. Why are such patterns common? Do they reflect an innate preference for uniformity, one that somehow guides change and is grounded in universal principles of language or psychology, or do they have another explanation? In linguistic theory the innate preference view has been reified under various names (Paradigm Coherence, Output–Output Correspondence, Uniform Exponence, etc.), but all such approaches posit distinct forces or constraints favoring uniformity.[2] The same view is also common in historical linguistics: "The motivation for leveling has been plausibly expressed in the slogan ONE MEANING—ONE FORM" (Hock and Joseph 1996: 155, following Hock 1991: 168).

I will argue, against this view, that there is no innate drive toward uniformity and that paradigm leveling is a by-product of independently motivated mechanisms of morphological change. Specifically, uniformity arises when the pattern of a non-alternating paradigm is imposed on a formerly alternating paradigm; it is in effect a type of extension. My argument is thus an "evolutionary" one in the sense of Blevins and Garrett (2004) and Blevins (2004a).[3]

The directionality questions in (3) are also important for our understanding of paradigm structure, and they relate to a second theme of this chapter: markedness. Leveling shows persistent regularities in directionality—singulars tend to replace plurals, nominatives tend to replace non-nominatives, third-person verb forms tend to replace first- and second-person forms, etc.—but there are exceptions and the reasons for the patterns remain unclear. Sometimes the patterns are associated with "markedness"—singulars are said to be unmarked *vis-à-vis* plurals, etc.—so if we

[1] I will not address a third crucial question, *selectivity*: why, under apparently similar circumstances, are alternations in one paradigm leveled while those in another paradigm are not?

[2] See e.g., Kiparsky (1972), Burzio (1996, 2000), Kenstowicz (1996), Buckley (1999), Steriade (1999, 2000), Benua (2000), and McCarthy (2005).

[3] The "evolution" metaphor has had other uses in language change (Haspelmath 1999a; Croft 2000; McMahon 2000), and I will generally avoid it. But analyses along broadly similar lines are available for a range of patterns: in syntax and morphosyntax, for split ergative case marking (Anderson 1977; Garrett 1990a), preposition–verb compounding (Garrett 1990b), adposition placement (Aristar 1991), and article–possessor complementarity (Haspelmath 1999b); in phonology, for dissimilation (Ohala 1981, 1993), velar palatalization (Guion 1998), metathesis (Blevins and Garrett 1998, 2004), compensatory lengthening (Kavitskaya 2002), antigemination (Blevins 2005a), positional vowel quality neutralization (Barnes 2006), consonant epenthesis (Blevins this volume), and other patterns (Blevins 2004a); in morphology, for infixation (Garrett 2001; Yu 2003) and templatic constructions (Good 2003); and in semantics and syntax, for numerous categories (e.g., Bybee 1988b; Bybee et al. 1994). Looming over the whole field is the work of scholars like Baudouin de Courtenay and Greenberg.

can understand directionality in leveling we may come to understand the basis of markedness.[4] This too bears on the theme of the present volume: if knowledge of some markedness patterns were innate, then directionality effects in leveling would be a manifestation of universal grammar. I will argue to the contrary that markedness patterns, as they emerge in paradigm leveling, are a product of the meaning and usage of the relevant categories.

I will address the questions in (2) and (3) through an analysis of changes affecting a well-defined part of the morphology of two languages, English and Ancient Greek. Specifically, I will look at changes in present vs. non-present verbal stem formation. The verbal systems of the two languages are organized along lines that are partly similar, partly different; the similarities and differences are both relevant. English has a basic opposition between *present* and *preterite* stems, which may be different (as in the *drive–drove* case above) or identical (as in standard *dive–dived*). Ancient Greek has an opposition between *present* (aspectually imperfective) and *aorist* (perfective) stems; a third basic category is the *perfect*. In both languages I will examine cases where alternations between present and non-present (English preterite, Greek aorist) stems were leveled, as well as cases where extension yielded alternations in previously non-alternating paradigms. My goal is to give a comprehensive picture of patterns of stem leveling and extension in the two languages.

The method I use is novel and yields new results. In studies of the optimization and directionality questions, the usual method is to analyze selected cases. These may be parade examples, such as the leveling of Latin rhotacism, or other cases where only an innate impulse toward uniformity seems to explain leveling. Such studies are essential, but the overall pattern of leveling and extension in a language also reveals generalizations that have not emerged from the study of individual cases.

In addressing the optimization question, the innate preference view of paradigm uniformity and the view I will sketch make different predictions about a language's overall dossier of levelings. I will argue that leveling is a special case of extension in which a non-alternating pattern is extended to a previously alternating paradigm. This requires that a suitable non-alternating model paradigm can be identified for every case of leveling. The innate preference view makes no such prediction; rather, since leveling is driven simply by a force favoring uniformity, it should be possible even without a non-alternating model paradigm. These predictions can be tested on a sufficiently rich dossier of examples. The facts turn out to contradict the innate preference view and to support the view that leveling is just extension.

In addressing the directionality question, my focus will be on a language-specific class of exceptions to an otherwise robust typological generalization. Exceptions to

[4] I share Haspelmath's (2006) reservations about the term "markedness", though obviously (contrary to his recommendation) I use the term here. What he calls "semantic markedness" is the specific type under discussion; in section 6.4 I also argue that there is a relation between formal and semantic markedness.

typological patterns, such as replacement of nominative by non-nominative forms, are usually assumed to be revealing (Tiersma 1982; Albright 2003, 2005, this volume). In this case the typologically widespread pattern is that presents are generalized at the expense of non-presents, and the systematic exception is in Ancient Greek. I will show that the Greek facts fit comfortably into no current theory of directionality but that they may be explained by a meaning-based account along lines first suggested by Kuryłowicz (1945–9).

The remainder of this chapter is organized as follows. I discuss the basic English and Ancient Greek data in sections 6.2 and 6.3 respectively. In section 6.4 I address the implications of the Greek data for the directionality question, and in section 6.5 I summarize and conclude. But first, I will very briefly sketch my assumptions about morphological change.

I assume that morphological production involves competition between the retrieval of memorized forms and the creation of new ones by rule, and that a mechanism of change is the creation of new forms if existing ones are not learned, remembered, or accessed fast enough (Bybee 1985; Barr 1994). This may have several causes. Learners may not hear an existing form, for instance, or they may hear it too infrequently to learn it. Alternatively, a morphological rule may be so salient that a new form is produced despite the existence of a memorized one; I will return to "salience" in section 6.4. In any case, sometimes an existing form is not reproduced and is replaced by a new form generated by rule. If this catches on, the older form may become otiose in a speech community and the newer form may replace it.

6.2 MIDDLE AND MODERN ENGLISH

English verbs are traditionally classified as *strong* or *weak*. Strong verbs have an unsuffixed preterite (i.e., past-tense form) and include various classes defined by their ablaut patterns; examples include *drive–drove* and *sing–sang*.[5] Weak verbs have suffixed preterites (e.g., with *-ed*) and fall into the three historical classes in (4), the last of which has several subtypes.

(4) English weak verb classes
  a. Regular weak verbs: no stem alternations (e.g., *like–liked*, *play–played*)
  b. *Rückumlaut* weak verbs: stem alternations due to earlier present-stem umlaut (e.g., *buy–bought*)[6]

[5] Here I mainly ignore the third principal part, the participle (e.g., *driven*, *sung*). An analysis of the role of participles would be essential for a full understanding of changes in English verb stem formation, but it would not affect my general argument.

[6] The term *Rückumlaut* ('back umlaut') alludes to the fact that umlaut affected present stem forms in these verbs.

c. 'Irregular' weak verbs:
   i. Stem alternations due to preterite-stem vowel shortening (e.g., *keep–kept*)
   ii. Stem alternations due to preterite suffix-vowel syncope (e.g., *hit–hit*)
   iii. Combinations and extensions of i–ii (e.g., *slide–slid*, *bend–bent*)

Various types of extension and leveling occur.[7]

The most frequent type of change, of course, is transfer from the strong class or an irregular weak class into the large class of regular weak verbs. Representative strong verbs that have undergone this change are *bake* and *help*, whose Old English preterites *bóc* and *healp* were replaced by *baked* and *helped* in the Middle English period. A representative irregular weak verb that has become regular is *reap*, whose seventeenth- and eighteenth-century preterite *reapt* [rɛpt] has been replaced by the regular weak form *reaped*.[8] All transfers of this frequent type have the effect of leveling—loss of a stem alternation—but can also be analyzed as extensions in which the morphological rule for regular weak verbs is applied to new items.

Rückumlaut weak verbs too have been transferred to the regular weak class. In Northern Middle English (Krygier 1997: 245–253), for example, *duell* 'dwell', *quell* 'kill', *quak* 'quake', and *wach* 'watch' show only regular weak preterites *duelled*, *queld*, *quaked*, and *wachit*, not the historically expected **dwald*, **cwald*, **quahte*, and **wahte*. Some other *Rückumlaut* verbs (e.g., *reche* 'reach', *teche* 'teach', *tell* 'tell') vary between regular weak forms (*reched*, *teched*, *teld*) and the historically expected forms (*raht*, *taht*, *tald*).[9] Here too transfer has the effect of leveling (*dwell–dwald* → *dwell–dwelled*), but it also amounts to extension of the non-alternating pattern of regular weak verbs.

Transfers into the irregular weak verb subclass in (4cii) could also be regarded as leveling. A representative case is *burst*, whose original strong preterite survives as Middle English *barst* but was replaced by *burst* beginning in the sixteenth century. (It is also transferred into the regular weak class as *bursted* or *busted*.) This is an extension of the pattern of *set*, *shut*, and other verbs properly belonging to the subclass in (4cii).

Transfers with the effect of leveling are common in the history of English verb inflection. But transfers also produce disuniformity. In Middle English (Marckwardt 1935), for example, the pattern of verbs like 3SG present *mynt* 'thinks', preterite *mynte* was extended to 3SG present *sent* 'sends', preterite *sende*, yielding a new 3SG preterite *sente* and other forms based on a new preterite stem *sent-*. From this point stem-final preterite devoicing was extended to the class of *gyrden* 'gird' (*girt*), *wenden* 'wend' (*went*), *spildan* 'spill' (*spilt*), etc., with stem-final sonorant + *d* clusters, and then to *leave–left*, *mean–meant*, etc.

[7] The changes discussed below are individually well known; for further information see the standard historical grammars (especially Jespersen 1942 and Luick 1914–40) and grammatical surveys (e.g., Mossé 1968). Note that the *þ* ('thorn') symbol writes an interdental fricative in Old English forms.

[8] Irregular weak verb inflection is itself an innovation for *reap*, originally a strong verb (Old English preterite singular *rap*).

[9] Still others (e.g., *buy–bought*) retain the *Rückumlaut* alternation.

A similar case is the creation of new irregular weak preterites formed on the model of *bleed–bled*, *chide–chid*, *greet–gret*, etc., with vowel "shortening" (quality change) in a stem ending in *d* or *t*. New present–preterite pairs of this type include *bite–bit*, *plead–pled*, *shoot–shot*, *slide–slid*, and *weed–wed*. In some cases the new form replaces a strong preterite (e.g., *slid*, replacing *slad*) and in others a regular weak preterite (e.g., *wed*, sporadic in Modern English, vs. regular *weeded*).

Finally, sometimes a verb is transferred from a weak class to a strong class. For example, *dig*, *sneak*, *stick*, and *string* were originally regular weak verbs with preterites *digged*, *sneaked*, *sticked*, and *stringed*; the strong preterites *dug*, *snuck*, *stuck*, and *strung* are first attested in Modern English.[10] Of course such changes cannot be classified as leveling (they produce non-uniform paradigms), and they always require an alternating model.

From this survey several important points emerge. First, as Bybee (1985: 51) notes, "changes in English verbs always involve a substitution of the Present base for the Past base." For example, the *reap–reapt* [rɛpt] alternation was not leveled by backforming a new present †*rep* but by forming a new preterite *reaped*, and the *dwell–dwald* alternation was not leveled by backforming a new present †*dwal*. I will return to the significance of this point in section 6.4.

Second, in every case, a change in preterite stem formation involved the transfer of a verb into a pre-existing class—in other words, extension of an existing pattern. Usually the non-alternating pattern of regular weak verbs was extended to formerly irregular verbs, but in other cases an existing strong or irregular weak pattern was extended to a formerly regular verb. The point is that each change can be treated as extension.

Third, in 900 years of Middle and Modern English linguistic history, there was never any case of pure leveling. The vast majority of preterite stem changes yielded paradigm uniformity, to be sure, but only given a pre-existing uniform paradigm of the same type. If a drive for uniformity were truly an independent force in language change, we would expect some cases in which uniformity is the sole motivation. For example, we might imagine that the Early Middle English paradigm of 'drive' could have been transformed as in (5), with present–preterite ablaut leveled.

(5) Early Middle English 'drive'
 a. Present 1SG *drīve*, 2SG *drīvest*, 3SG *drīveþ*, PL. *drīveþ*
 b. Preterite 1SG *draf*, 2SG *drive*, 3SG *draf*, PL. *driven*
 c. Hypothetical new preterite via leveling (present stem + preterite endings):
 1SG †*drīv*, 2SG †*drīve*, 3SG †*drīv*, PL †*drīven*

[10] At least for *dig* and *string*, interestingly, the strong forms were attested first as participles and only somewhat later as preterites. As for *sneak*, the earliest example of *snuck* in the online Oxford English Dictionary corpus (dictionary.oed.com) is from 1887 and most examples are American. The same corpus has almost twenty examples of preterite *sneaked* from before 1900, including two from the seventeenth century; the OED editors' assertion that *snuck* is the original preterite form is therefore false. (Other OED entry details can now also be corrected with data from the Literature Online corpus, lion.chadwyck.com. The verb *sneak* is first attested in John Studley's 1581 translation of Seneca's *Medea*, i.e., twenty years before the first example in OED s.v. *sneak*, from Shakespeare. The adjectival participle *sneaking* is first attested in a 1576 volume by George Whetstone, *The Rocke of Regarde*, fourteen years before the first example in OED s.v. *sneaking* ppl. a.).

The result of this hypothetical change would have been perfectly functional (e.g., 1SG present *drīve* vs. preterite †*drīv*), but no such change happened. The significant difference between this hypothetical change and the actual pattern of changes is that all the actual changes can be regarded as extension, in traditional terms, or the application of a morphological rule to new items. By contrast, in (5), no pre-existing paradigm existed—for example, a strong verb pattern lacking ablaut—that could have produced the new preterites. There simply was no basis for an extension.

An apparent case of leveling that should be noted because it is well known involves Verner's Law alternations in Old and Middle English. This is sometimes described as partial leveling, but it too is actually extension. A representative example is found in the Class II strong verbs, whose four Old English principal parts pattern as shown in (6a–c).

(6)		Infinitive	Past 1SG/3SG	Past plural	Participle	
	a.	bēodan	bēad	budon	boden	'command'
	b.	cēosan	cēas	curon	coren	'choose'
		sēoþan	sēaþ	sudon	soden	'boil'
	c.	abrēoþan	abrēaþ	abruþon	abroþen	'fail'

The example in (6a) is one of many verbs showing the regular ablaut pattern of Class II; those in (6b) are two of the six Class II strong verbs also showing consonant alternations, e.g., between *s* and *r* or *þ* and *d*, originally caused by Verner's Law.[11] The example in (6c) originally had Verner's Law consonant alternations as in (6b), but in Old English they had already been leveled, resulting in invariant root-final *þ*. This change also later affected some verbs like 'choose' in (6b); modern English retains ablaut (*ch**oo**se–ch**o**se*) but has eliminated the stem alternant with *r*. But this apparent case of leveling is easily explained as extension: the majority Class II pattern of *bēodan* in (6a), with ablaut but no consonant alternation, was extended to the minority with a consonant alternation.[12] Leveling is not an explanation or a driving force in this change.

More broadly, I take it that a theory that predicts the possibility of changes of a sort that never materialize in nearly a millennium of paradigmatic reshufflings—levelings that cannot be analyzed as extensions—is not a very good theory. I conclude that paradigm uniformity is not a force in change, even where it may result from change. Instead, uniformity arises by extension of a non-alternating pattern to previously alternating paradigms. In section 6.3 I will show that the same generalization holds for changes in Ancient Greek, but that Greek differs from English in the directionality of its present–non-present extensions and levelings.

[11] Verner's Law was an accentually conditioned fricative voicing change, transformed by later changes into alternations like those seen in (6b); the four other verbs like *cēosan* and *sēoþan* are *drēosan* 'fall', *forlēosan* 'lose', *frēosan* 'freeze', and *hrēosan* 'fall'.

[12] Similar accounts can be given of Verner's Law levelings in Classes I, III, and V.

6.3 ANCIENT GREEK

There are four major tense/aspect categories in Ancient Greek: the *present* (including also the imperfect), expressing imperfective aspect; the *aorist*, expressing perfective aspect; the *perfect* (including also the pluperfect and future perfect); and the *future*. Here I will examine changes in the formation of present vs. aorist stems; the perfect and future stems will play almost no role in the discussion. I will show, first, as shown for English in section 6.2, that while leveling leading to paradigm uniformity is common it can always be analyzed as extension and, second, that the directionality of paradigmatic changes in Ancient Greek is systematically different from that in English. I will suggest an analysis of the latter result in section 6.4.[13]

The remainder of this section is divided into five subsections: in section 6.3.1 I provide an overview of the relevant stem formation patterns; in section 6.3.2 I discuss changes affecting thematic presents in *-*y*-; in section 6.3.3 I discuss changes affecting nasal presents; in section 6.3.4 I discuss two apparent counter-examples to my claims about directionality; and I give a brief summary in section 6.3.5.

6.3.1 Present and aorist stem formation

Ancient Greek had numerous present stem types. They can be classified as *athematic* or *thematic* according to whether there is a theme vowel between the stem and ending; thematic and athematic presents also have different inflectional endings. For example, the athematic 1SG indicative ending is *-mi* (*deíknūmi* 'I show' from stem *deiknū-*) while its thematic counterpart is *-ō* (*basiléuō* 'I am king'). I will cite verbs by stem rather than as fully inflected forms, and I will cite the theme vowel as *-o*-; the present stem of 'be king' is thus *basileu-o-*.[14]

A summary of major athematic and thematic present stem classes is given in (7). The main athematic types are in (7a). Thematic types are listed in (7b), and include simple presents, nasal presents, and presents derived with a suffix *-*y*-. The *-*y*- suffix is lost phonologically due to prehistoric sound changes, with the three subtypes in (7biii).

(7) Major Ancient Greek present stem classes

a. Athematic: no theme vowel

i. Simple presents (e.g., *ei-mí* 'I am' from root *ei-*)

ii. Nasal presents, with suffix *-nā-* (*-nē-* in the Attic dialect) or *(n)nū-*

[13] Individually the changes discussed here are well known; for presentations of the data with references to additional literature see Schwyzer (1953) and Meier-Brügger (1992). I have derived my database of Greek changes from these handbooks and from other standard references cited by them.

[14] Citation by stem has the advantage of abstracting away from irrelevant morphophonemic changes. The theme vowel alternates between *-o-* (e.g., 1SG) and *-e-* (e.g., 2SG, 3SG); I write *-o-* for expository convenience.

b. Thematic: theme vowel before inflectional endings
   i. Simple presents
   ii. Nasal presents (suffix *-an-*, with nasal infix in light roots, i.e., roots ending in a short vowel and one consonant)
   iii. Suffix *-*y*- (always lost phonologically)
      α. Suffix *-*y*- preceded by a consonant (triggering various changes)
      β. 'Contract verbs' (suffix *-*y*- preceded by a short vowel)
      γ. Verbs in **-eu-y-* (denominative to noun stems in **-eu-*)

There are two main types of aorist formation: the non-sigmatic aorist, usually formed via root ablaut; and the sigmatic aorist, formed with a suffix *-s-* added to the present stem. (This suffix usually also has some effect on a preceding consonant.) The sigmatic aorists are productive, analogous to the English weak preterites; the asigmatic aorists are analogous to the English strong preterites. Aorists also have a prefix *e-* called the 'augment'. For example, the aorist stem of the verb 'be king' is *e-basileu-s-* (1SG *ebasíleusa*).

The largest present and aorist classes are the thematic presents and sigmatic aorists, but despite some subregularities there is no generally predictable relationship between the type of present stem a given verb has and the type of aorist stem it has. In virtually all present classes some verbs form sigmatic aorists, for example, but on the other hand some sigmatic aorists and some asigmatic aorists correspond to thematic presents.

6.3.2 Thematic presents in *-*y*-

The loss of *y with attendant effects on adjacent consonants led to opaque relationships between present and aorist stems in several subcategories of thematic present in *-*y*-. In turn a number of paradigmatic shifts can be attributed to these opaque relationships.

First, several transfers effectively shifted verbs into the class of vowel-final stems in (8), themselves a mix of historical *s*-stems (e.g., **teles-* 'finish') and long-vowel stems (e.g., *lū-o-* 'loose').

(8)

Aorist stem	Present stem	
e-tele-s-	tele-o-	'finish'
e-lū-s-	lū-o-	'loose'

For example, the class of historical *eu*-stem verbs developed presents in *-ei-o-* (< **-eu-y-*, i.e., **-ew-y-*) alternating with aorists in *-eu-s-*. As seen in (9), that alternation was leveled in favor of the aorist. (Arrows here and below indicate analogical rebuilding.)

(9)

Aorist stem	Earlier present stem	New present stem	
e-basileu-s-	*basilei-o-	→ basileu-o-	'be king'
e-douleu-s-	*doulei-o-	→ douleu-o-	'be a slave'
e-paideu-s-	*paidei-o-	→ paideu-o-	'be a child'

Likewise, as illustrated in (10), an alternation between long and short vowels in so-called "contract" verbs was leveled in favor of the aorist in various dialects of Greek.

(10)

Aorist stem	Earlier present stem	New present stem	
e-doulō-s-	doulo-o-	→ doulō-o-	'enslave'
e-poiē-s-	poie-o-	→ poiē-o-	'make, do'
e-step^{h}anō-s-	step^{h}ano-o-	→ step^{h}anō-o-	'crown'

Significantly, in neither case did leveling favor the present stem; for example, the aorist *e-basileu-s-* in (9) was not replaced by †*e-basilei-s-*, and the aorist *e-doulō-s-* in (10) was not replaced by †*e-doulo-s-*.

Second, Proto-Greek $^{*}k^{w(h)}$, $^{*}g^{w}$ underwent a conditioned split, creating alternations between aorist stems in *-ps-* (< **-k*^{w}*-s-*) and present stems in *-sso-* (< **-kyo-* < **-k*^{w}*-y-o-*) or *-zdo-* (from **-gyo-* < **-g*^{w}*-y-o-*). These alternations were leveled on the model of originally labial-final verbs (present stem in *-pto-* < **-p-y-o-*), like those in (11a); the result of leveling is shown in (11b).

(11) a.

Aorist stem	Present stem	
e-blap-s-	blap-t-o-	'injure'
e-kop-s-	kop-t-o-	'cut'

b.

Aorist stem	Earlier present stem	New present stem	
e-nip-s-	nizd-o-	→ nip-t-o-	'wash'
e-pep-s-	pess-o-	→ pep-t-o-	'cook'

Significantly, again, it was the aorist rather than the present stem that served as the base in leveling.

Two final examples show extension rather than leveling. First, because Proto-Greek **ky* > *tt* in the Attic dialect, thematic presents in **-y-o-* with roots ending in *k* show a regular alternation between aorist stems in *-ks-* and present stems in *-tt-o-*; see (12a). In Attic, as shown in (12b), the alternation in (12a) was extended to other verbs with aorist stems in *-ks-* from roots ending in *g*; they had originally had present stems in **-zd-o-* (< **-gy-o-*).[15]

(12) a.

Aorist stem	Present stem	
e-tarak-s-	taratt-o-	'disturb'
e-p^{h}ulak-s-	p^{h}ulatt-o-	'guard'
e-p^{h}armak-s-	p^{h}armatt-o-	'treat with drugs'

b.

Aorist stem	Earlier present stem	New present stem	
ēllak-s- (/e-allak-s-/)	*allazd-o-	allatt-o-	'(ex)change'
e-prak-s-	*prazd-o-	pratt-o-	'do'
e-sak-s-	*sazd-o-	satt-o-	'fill full'
e-sp^{h}ak-s-	sp^{h}azd-o-	sp^{h}att-o-	'slay'
e-tak-s-	*tazd-o-	tatt-o-	'put in order'

15 In the same dialect, leveling goes the other way in one case: the present stem *harpazd-o-* 'snatch away' spawned a new aorist stem *hērpas-* (/e-harpad-s-/) next to older *hērpaks-*. I cannot explain why this verb goes against the general pattern of the Attic dialect.

Second, because Proto-Greek **ty* > *tt* but **ts* > *s*, the Attic dialect had a pattern whereby aorist stems like *e-plas-* 'form, mould' corresponded to present stems like *platt-*. This pattern was extended to one verb, given in (13), with an original aorist stem in *s* and a present stem in *-zd-o-* (< *-*g*-*y*-*o*-).[16]

(13)	Aorist stem	Earlier present stem	New present stem	
	hērmos- (/e-harmo-s-/)	harmozd-o-	→ harmott-o-	'adapt, fit'

In the cases discussed in this section, note again that it is the aorist stem that serves as the base of leveling or extension, not the present stem, and that all examples of leveling can be analyzed as extension of a uniform pattern.

6.3.3 Nasal presents

I turn next to nasal presents: verbs whose present stems are marked by suffixes with a nasal consonant. They show several patterns of leveling. Shown in (14–15) are verbs with the suffix *-nā-* (= *-nē-* in the Attic dialect). Some verbs of this type acquired new presents in *-nnū-*, as shown in (14b), while a few others acquired presents in *-zd-*, as in (15b). It is crucial that the aorist stems of all types were marked simply by *-s*-; the pattern in (14a) was extended as in (14b), and the pattern in (15a) was extended as in (15b).

(14) a.	Aorist stem		Present stem		
	e-sbe-s-		sbe-nnū-	'extinguish'	
	e-zdō-s-		zdō-nnū-	'gird'	
	ēmp^{h}i-e-s- (/e-amp^{h}i-e-s-/)		amp^{h}i-e-nnū-	'clothe'	
b.	Aorist stem	Earlier present stem	New present stem		
	e-kera-s-	kir-nē-	→ kera-nnū-	'mix'	
	e-krema-s-	krim-nē-	→ krema-nnū-	'hang'	
	e-peta-s-	pit-nē-	→ peta-nnū-	'spread'	
	e-skeda-s-	skid-nē-	→ skeda-nnū-	'scatter'	

(15) a.	Aorist stem	Present stem		
	e-dika-s-	dika-zd-o-	'judge'	
	e-p^{h}ra-s-	p^{h}ra-zd-o-	'tell'	
b.	Aorist stem	Earlier present stem	New present stem	
	e-dama-s-	dam-nē-	→ dama-zd-o-	'tame'
	e-pela-s-	pil-nā-	→ pela-zd-o-	'approach'

Significantly, again, it was the aorist that served as a pivot. The aorist stem *e-peta-s-* in (14b) could imaginably have been replaced by †*e-pitnē-s-* based on the present stem, and *e-dama-s-* in (15b) could have been replaced by †*e-damnē-s-*, but this did not happen.

Verbs in *-n-C-an-* are illustrated in (16). Some details are obscure (Schwyzer 1953: 699–701), but the pattern is robust and clearly relies on an aorist pivot. We do not know for sure precisely which of the verbs in (16) were the source of the pattern, but evidently it was extended from examples where the nasal 'infix' belonged to the root.

[16] The aorist stem in (13) is underlyingly /e-harmod-s-/; cf. *harmódios* 'fitting'.

(16)	Aorist stem	Earlier present stem	New present stem	
	e-dak-	dakn-o-	→ daŋk-an-o-	'bite'
	ērug-	ereug-o-	→ eruŋg-an-o-	'discharge'
	e-t^{h}ig-	[none]	→ t^{h}iŋg-an-o-	'touch'
	e-lab-	lazd-o-	→ lamb-an-o-	'take'
	e-lat^h-	lēt^h-o-	→ lant^h-an-o-	'escape notice'
	e-lak^h-	[none]	→ laŋk^h-an-o-	'obtain'
	e-lip-	leip-o-	→ limp-an-o-	'leave'
	e-mat^h-	[none]	→ mant^h-an-o-	'learn'
	e-pat^h-	pask^h-o-	→ pant^h-an-o-	'suffer'
	e-put^h-	peut^h-o-	→ punt^h-an-o-	'inquire'
	e-p^{h}ug-	p^{h}eug-o-	→ p^{h}uŋg-an-o-	'flee'

6.3.4 Apparent counter-examples

Thus far in section 6.3, I have argued that Ancient Greek stem uniformity only arose in cases that can be analyzed as extension of an existing uniform pattern, and that the aorist was always the base of leveling. There are apparent counter-examples to this second claim: cases where the present serves or seems to serve as the base of leveling. Here I will treat two classes of counter-examples.[17]

The largest class of counter-examples is just comparable to those English strong verbs that are transferred into the regular weak class. A similar process is common in Greek: transfer from the class forming non-sigmatic aorist stems by root ablaut, into the class of sigmatic aorists. Two representative examples of this transfer are shown in (17).

(17)	Present stem	Earlier aorist stem	New aorist stem	
	ag-o-	ēgag- (/e-agag-/)	→ ēks- (/e-ag-s-/)	'lead'
	leip-o-	e-lip-	→ e-leip-s-	'leave'

The new aorist forms cited in (17) were at first sporadic but eventually became common or regular (Mandilaras 1973: 143–145). Such cases are clearly extension, with the present, not the aorist, serving as the pivot.[18]

To understand why this transfer type goes against the directionality pattern of other changes (where aorists serve as pivots), it is important to realize that the great majority of Ancient Greek verbs have thematic present stems and sigmatic aorists; an example is 'send' (present stem *pemp-o-*, aorist *e-pemp-s-*). Since classes with many members tend generally to draw other items into them, there was an independent pressure on

[17] For a third class of counter-examples I have no account. In some dialects other than Attic, the original *k–zd* alternation in (12b) was extended to new verbs, whose original aorist stems were replaced by stems in *-k-s-*. For example, in Cretan (Bile 1988: 219–221), based on present *psap^{h}izd-o-* 'count', aorist *e-psēp^{h}is-* was replaced by *e-psap^{h}ik-s-*.

[18] Note that the paradigm of 'leave' (present *leip-o-*, aorist *e-lip-*) generated both the new present in (16) and the new aorist in (17); the extension in (16) was earlier (attested in early Greek poetry) and the shift in (17) was later (attested in the comic poet Aristophanes and then only in much later sources).

verbs to fall into the class with thematic present stems and sigmatic aorists. Most of the transfers shown above—(9–10), (11b), (12b), (13), (15b)—are in fact of precisely this type. But in all those other cases, the initial paradigm had a sigmatic aorist on the basis of which a new present stem could be formed. The exceptional pattern in (17) involves verbs which had no sigmatic aorist to begin with.

In short, there is no way of transferring a root like 'lead' or 'leave' in (17) into the relevant large target class except by generating a new sigmatic aorist stem based on the present. Aorist stems like *ēgag-* and *e-lip-* themselves fall into no patterns that would generate a new present stem. The broader generalization is that the otherwise prevalent directionality patterns in a language fail just in case the only possible transfer demands an unexpected pivot.[19]

A second apparent counter-example to the generalization that Ancient Greek aorists rather than presents are pivots of extension and leveling involves Grassmann's Law. This was a sound change which had the effect of deaspirating the first of two successive aspirated stops in a root. For example, from a root $^*t^hrik^h$- 'hair', nominative singular t^hriks shows regular deaspiration ($^*k^hs > ks$) before nominative singular *-s* while genitive singular $trik^hós$ shows the effect of Grassmann's Law.

Grassmann's Law alternations were retained in nouns, as just indicated, and in some verbs, but they were leveled in other verb paradigms. As seen in (18), this affected the aorist and future stems of three verbs.

(18)	Present stem	Aorist stem	Future stem	
	peit^h-o-	*e-p^{h}eis- → e-peis-	*p^{h}eis- → peís-o-	'persuade'
	peut^h-o-	e-put^h-	*p^{h}eús- → peús-o-	'learn'
	teuk^h-o-	*e-t^{h}euk-s- → e-teuk-s-	*t^{h}eúks- → teúk-s-o-	'make'

Note that leveling seems to favor the present stem in each case.

In this apparently exceptional case, however, there is clear evidence that the pivot of the change was not in fact the present stem at all but another form elsewhere in the paradigm: the perfect middle. For the leveling in (18) happened only in those paradigms with a perfect middle whose forms also underwent Grassmann's law. This is shown in (19), where the perfect middle participles are given in the form they would have had prior to a set of consonant + *m* assimilations.

(19)	Present stem	Perfect middle participle	Aorist or future stem
	peit^h-	*pe-peit^h-ménos	*e-p^{h}eis- → e-peis-
	peut^h-	*pe-peut^h-menos	*p^{h}eus- → peus-
	teuk^h-	*te-teuk^h-ménos	*e-t^{h}euk-s- → e-teuk-s-

[19] For example, in Latin nominal paradigms, the usual directionality of leveling was (as is also common cross-linguistically) that the nominative influenced non-nominative forms. Against this generalization, the famous leveling of nominative singular *honōs* 'honor', genitive singular *honōris* (*honōs* → *honor*) was a transfer into the class of *soror* 'sister', genitive singular *sorōris*. The shifting class was small; the model class was large and full of frequent and productive members, and crucially the only point of contact between the two paradigms was outside the nominative. Similarly, in the Greek case, the point of contact between the two paradigms is outside the aorist.

Perfect middle forms like **pepeit*^{h}*ménos* (< * *p*^{h}*ep*^{h}*eit*^{h}*ménos*) in (19) also show the effects of Grassmann's Law.

By contrast, as seen in (20), there was no leveling of Grassmann's Law alternations in paradigms without a perfect middle.

(20) a. Present stem *ek*^{h}*-o-* 'have' < **hek*^{h}*-* with Grassmann's Law
Aorist stem *e-sk*^{h}*-*
Future stem *hek-s-o-*
Perfect middle: defective
b. Present stem *trek*^{h}*-o-* 'run' < **t*^{h}*rek*^{h}*-* with Grassmann's Law
Future stem: *t*^{h}*reks-o-*
Aorist stem: suppletive.
Perfect middle: suppletive.

There was also no leveling if the perfect middle underwent a *-p*$^{(h)}$*m-* > *-mm-* assimilation. This is shown by the example in (21).[20]

(21) Present stem *trep*^{h}*-o-* 'nourish' < **t*^{h}*rep*^{h}*-o-* with Grassmann's Law
Aorist stem *e-t*^{h}*rep-s-*
Future stem *t*^{h}*rep-s-o-*
Perfect middle participle *te-t*^{h}*rám-menos*

The significant contrast here is between the verbs in (20–21), which retained alternations created by Grassmann's Law, and those in (18–19), which leveled them. Between these two sets of verbs the only meaningful difference is that the verbs in (18–19) had perfect middle forms that underwent Grassmann's Law. In the literature on Greek historical morphology, it is well established that the perfect middle played a pivotal, influential role in the evolution of many patterns (Chantraine 1961); the case at hand is thus simply another example.

6.3.5 Summary

I now summarize the two main arguments of this section. First, present-aorist stem alternations were leveled in Ancient Greek only if a model uniform paradigm existed to serve as the analogical basis, never in contexts lacking such models. This is contrary to the prediction of the view that uniformity is a target of change, that a desire for uniform paradigms somehow drives language change. Second, against the characteristic pattern of English and many other languages, the aorist rather than the present stem regularly serve as the pivot in cases of extension and leveling. If directionality is an effect of "markedness", then the aorist rather than the present is the "unmarked" Ancient Greek aspectual category.

[20] I assume that the *-p*$^{(h)}$*m-* > *-mm-* assimilation in (21) preceded the changes that eventually happened in the perfect middle forms in (19): *-t*$^{(h)}$*m-* > *-sm-* and *-k*$^{(h)}$*m-* > *-ŋm-*. This is a reasonable assumption given that *-p*$^{(h)}$*m-* clusters survive nowhere, in any position, in any Greek dialect after 1200 BCE. In the earlier Mycenaean dialect, neither assimilation had happened.

6.4 DIRECTIONALITY AND THE ORIGIN OF MARKEDNESS

Why did a non-present form (the aorist) serve as the basis for leveling in Ancient Greek, when in English and other languages it is always the present that plays this role? Here a brief survey of theories of directionality may be useful. Many authors have discussed the phenomenon, and several proposals exist to account for the attested patterns. Three of these are essentially morphophonological. For example, Schindler (1974) argued that directionality patterns can be sensitive to the type of opacity of a phonological pattern being leveled or extended. Barr (1994) appealed to morphophonological idiosyncrasy in general; more recently, Albright (2003, 2005, this volume) has argued that the base in analogical change is, or is determined from, a surface form from which other surface forms can most effectively be predicted. None of these accounts based on morphophonological patterns readily explains the Ancient Greek data; Greek presents in particular belong to more morphological subpatterns than aorists, and thus presents are relatively hard to predict from the forms of their aorists.

The most common account of directionality effects appeals to frequency: the base is the most frequent form. This view was classically expressed by Verner (1875), Paul (1880), Wheeler (1887), and other authors, and is more recently associated with Bybee (e.g., 2001). Indeed, to explain the pattern in which the present is the base of analogical changes, it has been noted that presents are more common than non-presents. However, according to Duhoux (1992: 502–503), the text frequency of Ancient Greek tense/aspect forms is as shown in (22) for indicative and non-indicative verbs.[21]

(22)

Indicative (49% of total)	Non-indicative (51% of total)
Present + Imperfect: 55%	Present: 49%
Present: 23%	Aorist: 40%
Imperfect: 32%	Other: 10%
Aorist: 34%	
Other: 11%	

The base of paradigmatic extensions and leveling in Greek, the aorist, is therefore not the most frequent form (34–40 percent aorist vs. 49–55 percent present).

Another approach to directionality is that of Jerzy Kuryłowicz, who wrote that "[s]o-called 'analogical' actions follow the direction: *formes de fondation* → *formes fondées*, whose relationship emerges from their spheres of usage" (1945–9: 23). Later expanding on this idea, he wrote the following (Kuryłowicz 1964: 37–39):

In order that a proportion $a : b = c : d$ ... be valid and correct, the relation between a and b, c and d, must be shown to be a relation between basic and founded form ... [This

[21] Duhoux's statistics are based on earlier studies of a range of authors. Of course it can be objected that frequencies in surviving texts are not the same as the real-life frequencies that language users would have been exposed to; but the surviving texts do include colloquial as well as other texts and a difference of 10–15 percent is in any case not trivial.

relation] is due to the respective ranges of occurrence [= spheres of usage] of *a* and *b*, *a* being both neutral and negative, *b*, positive. This means that *a* is used also (as a neutral member) outside the opposition *a* (negative): *b* (positive). Such is the normal relation . . . e.g., between Latin *lupus* (neutral use = without distinction of sex; negative use = male): *lupa* (positive use = female).

On this approach, it would be necessary to say that the English present is the "neutral" member of the present–preterite opposition and the Ancient Greek aorist is the "neutral" member of the present–aorist opposition. Is there evidence to support this view? Does the Ancient Greek aorist have a broader "range of occurrence" or "sphere of usage" than the present?

I suggest that evidence of three types supports this view. First, in morphology, there are a few cases where a contrast between competing present forms is neutralized in the aorist. This is shown in (23).

(23)	Aorist	Present I	Present II ("durative")	
	e-sk^h-	ek^h-o- < *sek^h-o-	isk^h-o- < *si-sk^h-o-	'have, hold'
	e-mein-	men-o-	mi-mn-o-	'stay, stand fast'
	e-trō-s-	trō-o-	ti-trō-sk-o-	'kill, damage'
	e-kale-s-	kale-o-	ki-klēsk-o-	'call'

To be sure, relatively few verbs retain the archaic pattern shown here.

In semantics there is additional relevant evidence of two types. One concerns what is traditionally called "markedness" (what Haspelmath 2006 calls "semantic markedness", i.e., the phenomenon of interest to Kuryłowicz). In discussing the Russian perfective–imperfective aspectual contrast, Filip (2000: 82–83) comments as follows (cf. Jakobson 1932): "Since imperfective verbs can be used to denote total (or complete) events, that is, with the same function as perfective verbs, in traditional and structuralist Slavistics they are considered to be the unmarked member in the aspectual opposition."

The Ancient Greek aorist has likewise been said to be the "marked" member of the present–aorist opposition because, in the present, the imperfective–perfective contrast is "completely neutralized" (Duhoux 1992: 48). But this view is incorrect, since in suitable aspectual contexts, as illustrated in (24), aorists may refer to the here and now.

(24) a. ***éblapsas*** *m' . . . ent^{h}áde nûn* ***trépsas*** *apò teík^{h}eos*
hinder:AORIST:2SG me hither now **turn:AORIST:PTCPL** from wall
'You **hinder** me . . . in now **turning** (me) hither from the wall.' (*Iliad* 22.15–16)

b. *mēdén ti lían dusp^{h}oreîn* ***parḗinesa***
NEG at.all very.much be.angry:INF **advise:AORIST:1SG**
'I **advise** (you) not to be too angry.' (Euripides *Andromache* 1234)

c. *tí* ***et^{h}aúmasas****?*
WH **be.surprised:AORIST:2SG**
'Why **are** you **surprised**?' (Aristophanes *Clouds* 185)

Moreover, the aorist is not only possible in the here-and-now context of the present, but it is in fact "preferred when . . . the author shows no particular interest in the verbal action, and he renders it in the most banal possible way from the aspectual point of view" (Duhoux 1992: 378). That is, just as Kuryłowicz's Latin masculine *lupus* 'wolf' can refer to a female wolf or wolves in general, while feminine *lupa* cannot refer to a male wolf or wolves in general, so the aorist can be used in the here-and-now temporal context of the Ancient Greek present.

The final argument concerns negation. It is a well-known feature of Ancient Greek that while positive imperatives are present or aorist according to the aspectual context, negative ("prohibitive") imperatives are aorist, as illustrated in (25).

(25) **mḕ** *katà* *toùs nómous* ***dikásēte*** . . .
NEG according.to the laws **judge:AORIST:SBJV:2PL**
'**Do not judge** according to your laws . . . ' (Demosthenes 21.211)

In other words, the present–aorist contrast is regularly neutralized in this modal context.

A similar neutralization is seen in non-imperative negative contexts. As Gildersleeve (1900: 106) wrote, "[t]otal negation is expressed by the aorist." The infinitives in (26) are coordinated and generally comparable, for example, but the negated infinitive is aorist while the other is imperfect (the aspectual context is imperfective).

(26) *mēdèn* ***hamarteîn*** *esti* *theôn* *kaì* *pánta* ***katorthoûn***
NEG **err:AORIST:INF** be:3SG of.gods and all **do.right:INF**
'Not **to err** and **to do** all things **right** is for the gods.' (Demosthenes 18.289)

Two further examples contrasting negation with imperfects (past tenses of the present) and aorists appear in (27a–b); both translations are by Gildersleeve (1900: 95).

(27) a. *hoi mèn* *ouk* ***êlthon***,
the PTCL NEG **come:AORIST:3PL**
hoi d' *elthóntes* *oudèn* ***epoíoun***
the PTCL come:AORIST:PARTICIPLE NEG **do:IMPF:3PL**
'Some **did** not **come**, and those who did come **would** not **do** anything.'
(Demosthenes 18.151)

b. *ouk* ***anébain'*** *epì tḕn naûn*
NEG **go.up:IMPF:3SG** on the ship
'He **would** not **go on board** the ship [as was expected].' (Demosthenes 21.163)

The meaning of the negated imperfect in both examples is that an event continued not to take place, rather than that it did not take place (or did not continue to take place). As schematized in (28), imperfective aspect has scope over negation while simple negation is expressed with the aorist.

(28) a. 'did not come' in (27a) = NOT (come)
b. 'would not board the ship' in (27b)
= IMPERFECTIVE (NOT (board the ship))

In short, semantic (and some morphological) evidence suggests that the Ancient Greek aorist–present relationship is semantically monotonic—in other words, that imperfective-aspect (present, imperfect) forms add a component of meaning to perfective-aspect (aorist) forms. I suggest that this semantic monotonicity lies behind the pattern of directionality seen in the extension and leveling of Greek stem alternations. Of course this in turn requires explanation—an appeal to "markedness" would beg the question—and two approaches suggest themselves. One would appeal to universal (or innate) preferences. For example, a basic preference for semantically monotonic morphological derivations might guide word formation and so attested patterns of change.

An alternative view within usage-based models of morphological change (Bybee 1985, 2001; Barr 1994) might appeal to salience as a crucial factor determining patterns of change. Assuming (as in section 6.1) that new forms arise when existing forms are not learned, remembered, or accessed fast enough, a form should be more vulnerable to replacement if it is less salient in memory and so less readily accessed than one derived by the morphological system. Known causes of salience include morphophonological irregularity (Paul 1880; Barr 1994) and high token frequency. We may hypothesize, likewise, that morphological categories with a broader sphere of usage (or less complex meaning) are more salient in memory, hence more easily accessed in language production, and hence serve as bases in the derivation of new forms.[22]

The choice between these two approaches, both of them obviously speculative and sketchy as stated here, returns us to the main theme of this volume. To what extent are patterns of change themselves consequences of built-in linguistic (or psychological) preferences? To what extent do they simply reflect the interaction of independent mechanisms? For the case at hand—directionality effects in extension and leveling—the answer is not yet known, but I have suggested that it will emerge from an understanding of the relation between morphology and semantics in language change.

6.5 CONCLUSION

I have made two main arguments in this chapter. First, I have shown that pure leveling does not exist and that the emergence of paradigm uniformity is always the imposition of an existing (uniform) pattern on a non-uniform paradigm. I conclude that paradigm uniformity is not an independent force or target in language change, for if a preference for uniform paradigms were an independent force we should see clear evidence for it somewhere among the almost limitless variety of attested changes. This in turn

[22] Susanne Gahl calls my attention to psychological studies (e.g., Hino and Lupker 1996) showing that some lexical decision and naming tasks are done faster for polysemous words than monosemous words; it is possible that this effect is related to the sphere-of-usage patterns under discussion.

bears on the general theme of this volume: paradigm uniformity, a common pattern in language, is diachronically epiphenomenal and not somehow embedded in universal grammar.

Second, I have identified a systematic difference between English and Ancient Greek in the directionality of paradigmatic changes. In English (and other languages), present-tense verb forms influence preterites; in Ancient Greek, presents are influenced by non-presents (aorists). I argued that this finding is not readily accommodated by theories invoking frequency or form predictability as the major factors influencing paradigmatic directionality, and that we need a more complex theory that also takes account of the semantics of morphological categories. Formulating such a theory more precisely and modeling the interaction of all of the factors that contribute to morphological change remain exciting projects for the future.

7

Explaining Universal Tendencies and Language Particulars in Analogical Change

Adam Albright
MIT

7.1 INTRODUCTION

It is well known that members of morphological paradigms exert an influence over one another, and forms are occasionally rebuilt to create more coherent and consistent paradigms. For example, in the transition from Middle High German (MHG) to New High German (NHG), the singular forms of verbs with *eu* ∼ *ie* alternations (strong class II) were rebuilt to contain *ie* throughout, as in (1) (Paul, Wiehl, and Grosse 1989: §242).[1]

(1) Loss of /iu/∼/ie/ alternations[2] in early New High German

'to fly'	MHG		Early NHG		NHG
1SG	*vl**iu**ge*	>	*fl**eu**ge*	⇝	*fliege*
2SG	*vl**iu**gest*	>	*fl**eu**gst*	⇝	*fliegst*
3SG	*vl**iu**get*	>	*fl**eu**gt*	⇝	*fliegt*

[1] I use the following orthographic conventions: X∼Y represents synchronic alternations between X and Y within a paradigm; X→Y represents a synchronic morphological or phonological rule changing input X to surface Y; X>Y indicates regular sound change from X to its expected outcome Y, while X⇝Y indicates that form X has been replaced by an analogically rebuilt form Y. Analogically rebuilt forms are also underlined in tables, to highlight those parts of the paradigm that have undergone changes. In all of the cases discussed here, the term PARADIGM refers to the set of inflected forms which share a single lexical stem (the set of case forms of a noun, the various person, tense, and number inflections of a particular verb, etc.).

[2] This alternation was produced by regular sound changes affecting the Proto-Germanic diphthong **eu*. Specifically, when *eu* preceded a syllable containing a high vowel, it raised to *iu*, and otherwise it lowered to *eo* and subsequently dissimilated to *io* > iə. Since the present singular suffixes all had high vowels (*-u*, *-is*, *-it*) and the plural suffixes all had non-high vowels (*-eːm*, *-et*, *-ant*), this resulted in singular ∼ plural alternations (Paul, Wiehl, and Grosse 1989: §35).

'to fly'	MHG		Early NHG		NHG
1PL	*vliegen*	>	*fliegen*	>	*fliegen*
2PL	*vlieget*	>	*fliegt*	>	*fliegt*
3PL	*vliegen*	>	*fliegen*	>	*fliegen*

In other verbs, singular∼plural vowel alternations were not lost, but were simply rearranged. For example, in strong verbs like *hëlfen* 'help', *nëmen* 'take', and *geben* 'give' (classes IIIb, IV, and V, respectively) the first-person singular form was rebuilt to match the plural, as in (2a) (Paul, Wiehl, and Grosse, 1989: §242). The result was a new alternation within the present tense paradigm, parallel to a separate pattern of alternation known as *umlaut*, seen in the verb *graben* 'dig' (2b).

(2) Rearrangement of /i/∼/ë/ alternations in early New High German

a. 'to give'	MHG	Early NHG		NHG	b. Following pattern of 'to dig'
1SG	*gibe*	*gibe*	⇝	*gebe*	*grabe*
2SG	*gibest*	*gibst*		*gibst*	*gräbst*
3SG	*gibet*	*gibt*		*gibt*	*gräbt*
1PL	*gëben*	*geben*		*geben*	*graben*
2PL	*gëbet*	*gebt*		*gebt*	*grabt*
3PL	*gëben*	*geben*		*geben*	*graben*

Although the changes in (1) and (2) yielded different patterns of alternation, what they have in common is that some members of the paradigm have been rebuilt to match other forms (*paradigm leveling*) or to differ systematically from another form (*analogical extension*, or *polarization*; Kiparsky 1968). The form that determines the shape of the rebuilt paradigm is traditionally referred to as the *base*, or *pivot*, of the change.

A long-standing issue in the study of analogy is the question of which forms act as bases and which are rebuilt. Typically, this is cast as a typological question: are there certain forms that tend to serve as bases, and other forms that tend to be rebuilt? Careful inspection of many cases has revealed numerous tendencies: analogy tends to be based on frequent forms, shorter forms, morphosyntactically less marked forms, and so on (Kuryłowicz 1945–9/1995; Mańczak 1958; Bybee 1985; Hock 1991). These tendencies are often taken to be primitives of historical change: change can eliminate alternations by replacing less frequent alternants with more frequent ones, marked forms with unmarked ones, and so on.

Much less attention has been devoted to explaining the language-particular aspects of analogy. Why does analogical change favor a particular base in a particular language? Why are alternations sometimes leveled, and sometimes extended? The typological approach makes only weak predictions about individual cases: certain changes are universally more or less likely, and the fact that a particular language underwent a particular change is a statistical accident. To the extent that individual changes obey

the typological tendencies, they can be seen as reasonable and natural, but analyses of analogical change seldom commit to the claim that an attested change was the only analogy that could possibly have occurred.

In Albright (2002a), a model is proposed that makes precisely this claim. Specifically, it is hypothesized that learners select base forms as part of a strategy to develop grammars that can produce inflected forms as *reliably* or as *confidently* as possible. In order to do this, learners compare different members of the paradigm, using each to attempt to predict the remainder of the paradigm with a grammar of stochastic rules. The part of the paradigm that contains as much information as possible about how to inflect the remaining forms is then selected as the base form, and a grammar is constructed to derive the rest of the paradigm. In this model, analogical change occurs when the resulting grammar derives the incorrect output for certain derived (non-basic) forms, and these errors come to replace the older, exceptional forms. Thus, all analogical change is viewed as (over)regularization, echoing earlier proposals by Kiparsky (1978) and others. Since the procedures for base selection and grammar induction are both deterministic, this model makes strong predictions about possible analogical changes: they must be based on the most informative form in the paradigm, and the only possible "analogical" errors are those that can be produced by the rules of the grammar. I demonstrate that these predictions are correct in several typologically unusual cases.

The goal of this chapter is to show that a confidence-based model can make correct predictions not only about individual cases, but also about the typology of analogical change. It is organized as follows: first, I provide a brief overview of tendency-based vs. structurally based approaches to analogical change, summarizing the major generalizations that have been uncovered, and situating the current work in an area that has been approached from radically different perspectives. Next, I present an overview of the synchronic model developed by Albright (2002a). I show first how the synchronic confidence-based approach can explain the direction of analogy in individual cases, and then move on to explore its typological implications. I consider first some apparent counter-examples to the confidence-based approach, showing that in at least some cases where analogy has seemingly favored an uninformative member of the paradigm, that form is not nearly as uninformative as it might appear. In other words, many apparently unusual cases are not as surprising as they would seem based on schematic presentations. I then consider how token frequency may affect the calculation of confidence, favoring the selection of more frequent forms as bases. An exploration of the parameter space of the model reveals that even without an explicit bias to select more frequent forms, they are nonetheless selected as bases under most conditions. Thus, a model of grammar induction that aims to construct accurate and reliable grammars is able to derive the observed typological tendencies without any built-in bias specifically designed to favor more frequent forms as bases.

7.2 TWO APPROACHES TO ANALOGICAL CHANGE

7.2.1 A typological approach

Starting with the neo-grammarians, formal analyses of language change have divided changes into two types. On the one hand, there is phonetic sound change, which is said to be regular and law-abiding, in the sense that it is (in principle) exceptionless and can be described formally as the operation of rules. On the other hand, there are non-phonetic changes, such as analogy and reanalysis, which are claimed to be sporadic, unpredictable, and describable only by tendencies, not laws.

A consequence of this division is that there is a sharp difference in how universals and the relation between synchrony and diachrony have been approached in the two domains. For phonetic change, it is generally accepted that there is a close relationship between synchrony and diachrony, even if the exact nature of the relationship is debated. Diachronic change creates synchronic alternations, and synchronic considerations such as articulation and perception motivate diachronic change. Furthermore, since phonetic pressures are at least in some respects universal, it is intuitively clear that some sound changes should be more common or natural than others. As a result, changes should go in some directions but not others. Explanation consists of uncovering the universal pressures, determining whether they are diachronic or synchronic in nature, and modeling them with the most restrictive possible theory that captures both the language-particular and typological patterns.

For non-phonetic changes like analogy, on the other hand, the situation is quite different. Whereas phonetic change is rooted to a large extent in physical and perceptual pressures, analogy is driven by more abstract cognitive considerations, such as reducing alternations within paradigms, or reducing the number of patterns in the language. Traditionally, these pressures have held no formal status in synchronic grammar,[3] but are seen as a diachronic force or acquisition bias that gradually eliminates alternations and restores regularity. Since these pressures do not place any restrictions on *how* regularity is achieved, analogical change may proceed in many different directions, and it is often difficult to classify one change as more or less natural than another competing possibility. Continuing with the example from above, in Middle High German, many verbs exhibited vowel alternations between the plural and some or all of the singular, as in (3a). In Modern German, these alternations have been retained in some cases (e.g., 'know' (3b.i) and most of the modal verbs), lost in others (e.g., 'fly' (3b.ii)); in yet other verbs such as 'give' (3b.iii), the alternation was retained in just some forms (the second- and third-person singular), as shown in (2) above.

[3] A recent exception is the reliance on PARADIGM UNIFORMITY or UNIFORM EXPONENCE constraints in Optimality Theory (Burzio 1994; Kenstowicz 1996; Steriade 2000; McCarthy 2005).

(3) Paradigmatic changes in early New High German

a. Alternations in Middle High German present tense paradigms

	i. 'know'	ii. 'fly'	iii. 'give'
1SG	*weiʒ*	*vliuge*	*gibe*
2SG	*weist*	*vliugest*	*gibest*
3SG	*weiʒ*	*vliuget*	*gibet*
1PL	*wiʒʒen*	*vliegen*	*gëben*
2PL	*wiʒʒet*	*vlieget*	*gëbet*
3PL	*wiʒʒen*	*vliegen*	*gëben*

b. Modern German paradigms (analogically changed forms are underlined)

	i. 'know'	ii. 'fly'	iii. 'give'
1SG	*weiß*	*fliege*	*gebe*
2SG	*weißt*	*fliegst*	*gibst*
3SG	*weiß*	*fliegt*	*gibt*
1PL	*wissen*	*fliegen*	*geben*
2PL	*wisst*	*fliegt*	*gebt*
3PL	*wissen*	*fliegen*	*geben*

The change from (3a) to (3b) represents a modest simplification or regularization: singular~plural alternations have mostly been eliminated except in a few high-frequency verbs (such as *wissen*), leaving just two general patterns (non-alternation, and raising in the second- and third-person singular). Logically, there are many other possibilities that seem just as natural, however. Could analogical change have gone further, eliminating alternations in all verbs? Or could it have gone in a different direction, yielding paradigms like *fleuge*, *fleugst*, *fleugt*, *fleugen*, *fleugt*, *fleugen*, or perhaps *fliege*, *fleugst*, *fleugt*, *fliegen*, *fliegt*, *fliegen*?

Under the traditional view of analogy, the answer is affirmative: changes in any direction are possible. Nevertheless, it is commonly accepted that some changes are more likely than others. Analogical changes are often based on the shortest or least suffixed member of the paradigm (Mańczak 1958; Bybee 1985: 50–52; Hayes 1995), the least marked member of the paradigm (Jakobson 1939; Greenberg 1966a; Bybee and Brewer 1980; Tiersma 1982; Bybee 1985), and the member of the paradigm with highest token frequency (Mańczak 1980: 284–285). In many cases, all three of these factors converge, yielding a base form that is frequent, unmarked, and unsuffixed (such as a nominative singular, or a third-person singular present form). At the same time, there are many cases in which these factors do not converge, and a subsequent change obeys one trend at the expense of others. Even more troubling, there are analogical changes that apparently violate all of these tendencies, rebuilding paradigms on the basis of a less frequent, more marked, suffixed base form (Hock 1991: ch. 10).

A well-known example of analogy based on a marked form involves the loss of final devoicing in Yiddish (Sapir 1915: 237; Kiparsky 1968: 177; Vennemann 1972a: 188–189; Sadock 1973; King 1980). In its earliest stages, Yiddish, like Middle High German, had

final devoicing of obstruents (seen here in alternation between singular *vek* and plural *vegə* in (4a)). However, in many dialects of Yiddish, final devoicing was subsequently lost, and the voicing value of the plural was reintroduced to the singular, leaving paradigms with [g] throughout as in (4b). (The change of the plural suffix from Ø to -ə *n* is irrelevant for the point at hand.)

(4) Loss of final devoicing in Yiddish

'way'	a. MHG			b. Early Yiddish			c. Modern Yiddish	
	SG	PL		SG	PL		SG	PL
NOM, ACC	*vek*	*vegə*	>	*vek*	*veg*(ə)	⇝	*veg*	*vegən*
GEN	*vegəs*	*vegə*	>	*vegəs*	*veg*(ə)			
DAT	*vegə*	*vegən*	>	*veg*(ə)	*vegən*			

We can confirm that the change from *vek* to *veg* is due to the voicing in the plural form, and not to a separate process of final voicing, by comparing voiceless-final stems and observing that they remain voiceless (5).

(5) Voiceless-final stems remain voiceless

'sack'	a. MHG			b. Earlier Yiddish			c. Modern Yiddish	
	SG	PL		SG	PL		SG	PL
NOM, ACC	*zak*	*zekə*	>	*zak*	*zek*(ə)	⇝	*zak*	*zek*
GEN	*zakəs*	*zekə*	>	*zakəs*	*zek*(ə)			
DAT	*zakə*	*zekən*	>	*zak*	*zekən*			

As Sapir first noted, this change is a paradigmatic one: words with [g] in the plural generally had [g] restored in the singular as well, while words with no plural form (such as the adverb *vek* 'away') did not change.[4] Thus, it appears that in this case, paradigms have been leveled to the form found in the plural, in spite of the fact that plurals are more marked and less frequent than singular forms.

Cases of leveling to marked forms are not uncommon in the literature, and the usual response has been to claim that the direction of analogy reflects general typological tendencies, but is not governed by any hard and fast rules (Kuryłowicz 1945–9; Mańczak 1958; Hock 1991). This position is summed up succinctly by Bybee and Brewer (1980: 215):

> A hypothesis formulated in such a way makes predictions of statistical tendencies in diachronic change, language acquisition and psycholinguistic experimentation. It cannot, nor is it intended to, generate a unique grammar for a body of linguistic data.

Although this is a reasonable approach to finding and testing descriptive hypotheses, it is unsatisfying from an explanatory point of view. As an account of language change, it tells us what changes are likely in general, but it cannot tell us why a particular language changed in a particular way at a particular time. As an account of language acquisition or experimental results, it tells us what types of errors or

[4] The details of the loss of final devoicing are considerably more complex than what is described here; see King (1980) and Albright (to appear) for an overview.

results we might expect of humans in general, but it cannot explain why speakers of different languages behave differently in the types of errors they make or in their responses to psycholinguistic experiments. In order to explain data from speakers of individual languages, we need a synchronic, language-particular understanding of how paradigms are organized, and how this organization determines what is a possible analogy.

7.2.2 A grammar-based approach

It has been noted at least as far back as Hermann Paul that analogical change is quite plausibly rooted in the way that children learn language, and the (sometimes incorrect) analyses that they impose on it (Paul 1920: ch. 5). Early generative approaches to language change attempted to formalize this intuition, proposing that analogy might be best explained not by examining the surface patterns before and after the change, but rather by comparing the change in the underlying grammatical analysis. Kiparsky (1968) and King (1969: ch. 5.3) advanced the hypothesis that analogical change serves to simplify grammar, either by removing rules from the grammar (rule loss), simplifying their environments (broadening, or generalization), eliminating opacity (maximizing transparency), or removing exceptions. For example, the loss of final devoicing in Yiddish may have involved analogy to a marked base form on the surface, but the resulting grammar was simpler: the final devoicing rule, which had become opaque due to the counterfeeding apocope rule, was lost. The intuition is that analogical changes have a *structural* motivation, which can be viewed perhaps as a learning bias for transparent rule orderings, or for grammars that use as few rules as possible.

The idea that analogy always results in grammar simplification is tantalizing, but unfortunately, there are many changes that cannot be straightforwardly analyzed as simplification. Naturally, whether or not a particular change can be viewed as simplification depends crucially on the grammatical analyses employed, but in many cases the resulting grammar is not obviously any simpler than the original one—see, e.g., King (1969) and Vennemann (1972a) for discussion. To take just one example, the change in the first-person singular of the verb *geben* in German from *gibe* to *gebe* ((2) above) does not eliminate the need for an *e*∼*i* vowel alternation rule, or simplify its environment—in fact it arguably makes it more complex, by trading in a singular/plural alternation for an alternation between the second- and third-person singular and all remaining forms. Thus, it is not possible to maintain a strong version of the hypothesis that analogy is always grammatical simplification.

Nonetheless, as Vennemann points out, there are structural considerations other than simplification that could play a role in determining the direction of analogy. Returning to the example of the loss of final devoicing in Yiddish, let us compare the actual outcome ((4) above) with a hypothetical version of Yiddish with leveling to the singular form.

(6) Hypothetical extension of final devoicing in Yiddish

'way'	a. Earlier Yiddish			b. Hypothetical development	
	SG	PL		SG	PL
NOM, ACC	*vek*	*veg*	⇝	*vek*	*vek*(*ən*)

Vennemann notes that although the hypothetical analogy in (6) employs an unmarked base, it would actually represent a rather unusual change. The reason is that analogy to the singular would eliminate the phonemic opposition in final position between underlying /k/ (that is, words like *zak* with [k] in both singular and plural) and underlying /g/ (like *veg*, with [k] ~ [g] alternations). He hypothesizes that analogy characteristically preserves phonemic distinctions: "Sound change neutralizes contrasts, analogy emphasizes contrasts by generalizing them" (Vennemann 1972a: 189). Vennemann calls this the *predictability principle*, and he suggests that the desire to maintain contrasts can override the tendency to level to unmarked base forms.

In Albright (2002a), it is proposed that the urge to maintain contrasts is more than a mere tendency that influences the direction of analogy; in fact, it forms the basis of how learners approach the problem of learning paradigms. The premise of this approach is that speakers ideally need to be able to correctly understand and produce inflected forms of their language. In order to do this, they cannot wait around to hear and memorize all forms of all words, since there are many forms that they will simply never encounter, particularly in a highly inflected language. Thus, learners need to make inferences about the phonological and morphological properties of words based on incomplete information. The proposal is that learners adopt a strategy of focusing on the part of the paradigm that contains the most contrastive information, and allows them to project the remaining forms as accurately or as confidently as possibly—that is, the most informative or predictive part of the paradigm. This form is chosen as the base of the paradigm, and a grammar is constructed to derive the remaining forms.[5] I will refer to this strategy as CONFIDENCE MAXIMIZATION, since its goal is to allow the learner to infer properties of words as confidently as possible.

On the face of it, a confidence-based approach appears to suffer from as many counter-examples as any of the tendencies discussed above. For example, a famous analogy in the history of Latin eliminated a stem-final contrast between [r] and [s]: *honōs* ~ *honōris* 'honor-NOM/GEN' ⇝ *honor* ~ *honōris*, on analogy with underlying /r/ in words like *soror* ~ *sorōris* 'sister-NOM/GEN' (for discussion, see Hock 1991: 179-190; Barr 1994: 509–544; Kenstowicz 1996; Kiparsky 1997; Albright 2005; and many others). Such cases are not necessarily a problem if predictability is viewed as just one more factor that can compete or conspire to determine the direction of analogy, but they are a challenge to the idea that bases are *always* the most predictive form.

[5] This search for contrastive information is similar to the way in which generative phonologists usually assume that underlying forms are discovered; for discussion of the parallels and differences, see Albright (2002a).

Why does analogy sometimes wipe out distinctions, if bases are always chosen to be maximally informative?

I hypothesize that the reason why analogy sometimes eliminates contrasts is that learners are restricted in the way that bases are chosen, and cannot always select a form that maintains all of the contrasts that are displayed in their language. In particular, I propose that there is a SINGLE SURFACE BASE restriction: learners must choose a surface form as the base, and the choice of base is global (that is, the same for all lexical items). When there is no single form in the paradigm that preserves all distinctions for all lexical items, the learner must choose the form that maintains distinctions for as many lexical items as possible.[6] In the case of Latin, the contrast between [r] (*soror*) vs. [s] (*honōs*) was neutralized to [r] in oblique forms (*sororis*, *honōris*). This neutralization affected relatively few forms, however, compared to neutralizations in the nominative caused by cluster simplification and morphological syncretism; thus, the globally best choice of base form in Latin would have been an oblique form, even if it neutralized the rhotacism contrast. Thus, we see that the single surface base restriction can result in certain contrasts being lost in the base form; when this happens, they will be open to analogical leveling. For example, in Latin, the minority [r] ∼ [s] alternation merged with the more regular [r] ∼ [r] pattern. Leveling under this approach is not a grammatical simplification, but rather lexical simplification, eliminating exceptions and replacing them with grammatically preferable regular forms.

It should be emphasized that nothing in this system requires that overregularization/leveling must take place; as long as learners have sufficient access to input data, there is always the potential to learn and maintain irregularity in derived (= non-basic) forms. The model does not make specific predictions about when leveling will occur, except that we would obviously expect it when input data about exceptional forms is reduced, such as in low-frequency words, reduced input because of bilingualism or language death, etc.; see also Kuryłowicz (1945–9/1995), Bybee (1985), Barr (1994), and Garrett (this volume) on this point. In fact, I am largely in agreement with Garrett's claim that the morphological change is driven diachronically by inaccurate or incomplete transmission of the full set of inherited forms. The current model differs from his account, however, in positing a cognitive constraint on the form of possible grammars: namely, that all morphological rules refer to the same base form as their input. This restriction is what allows the model to make strong predictions about directionality: when change occurs, it should always involve replacing an exceptional non-basic form with an innovative regularized form. Although this restriction is not

[6] All of the cases discussed here involve a single base form within a rather limited local paradigm (one tense of a verb, singular and plural forms of a noun, etc.). An important question not addressed here is whether larger paradigms, with multiple tenses, aspects, etc., might involve multiple local bases—perhaps along the lines of the traditional principal parts analysis of Latin or Greek verbs. The question of what considerations might compel learners to establish multiple base forms is a matter of ongoing research; some examples are discussed in Albright (2002a: §6.3).

a logically necessary part of the formalism, it receives empirical justification from the fact that analogy is overwhelmingly unidirectional, to a far greater extent than token frequency, memory failings, and chance would predict. In Classical Latin, nominatives were rebuilt on the basis of an oblique form, while in English, preterite forms were always rebuilt on the basis of presents (Garrett, this volume) and in Ancient Greek, presents were rebuilt on aorists (Garrett, this volume), and so on.[7] I take such asymmetries to be the fundamental explicandum of analogical change.

Unlike the tendency-based approach, the confidence-based explanation of analogy aims to capture the directionality of *all* cases of paradigmatic change—a steep task, given the typological diversity of attested changes. In Albright (2002a), I examined several typologically unusual analogies, showing that they were indeed based on the most predictive or informative member of the paradigm. In order to demonstrate the validity of this approach in general, however, two questions must be answered. The first is a question of coverage: are all analogical changes really based on the most predictive member of the paradigm, or are there cases in which analogy favors a less predictive member of the paradigm, contrary to the predictions of confidence maximization? The first goal of this chapter is to examine one class of apparent counterexamples, using data from an analogical change currently underway in Korean. I show that even though a form may appear to be radically uninformative when viewed schematically, it may actually be the most predictive form when we consider the language as a whole. Thus, some apparently exceptional changes need not be seen as counterexamples at all, once we have a suitable understanding of the factors that play a role in the calculation of confidence.

The second question that must be answered is a typological one: can a confidence-based approach explain why certain types of analogy are extremely common, while others are relatively rare? In general, such questions play a secondary role in structural analyses; as long as the predicted patterns are a good match to the attested patterns, the question of why some are chosen more frequently is often ignored (though see, e.g., Harris (this volume) for one possible line of explanation). Nonetheless, the fact remains that analogy is very often based on unmarked or frequent members of the paradigm, and this fact demands some sort of explanation. The second goal of this chapter, then, is to explore the behavior of the model when exposed to a large range of possible input languages. I will show that because of the way that confidence is calculated, the model does indeed select bases with high token frequency a majority of the time; less frequent forms are chosen as bases only under somewhat extreme conditions, mirroring the observed typology.

Before we can answer these typological questions, however, it is first necessary to provide an overview of the synchronic model of base selection and grammar acquisition.

[7] See Garrett's treatment of apparent counter-examples in Greek (this volume), and also Tiersma (1982) and Bybee (1985), for discussion of factors that might reverse the ordinary directionality of a change.

7.3 A SYNCHRONIC MODEL OF PARADIGM ACQUISITION

For languages with even a modest amount of morphological complexity, learners face a number of difficult tasks in learning to accurately understand and produce paradigms of inflected forms. They must learn what inflectional categories are marked in the language, what the relevant markers are, which words belong to which morphological classes, and which words exhibit irregularities or alternations that cannot be predicted by rule. Furthermore, if there is neutralizing phonology (as is often the case), the learner must be able to compare related forms, determining where in the paradigm one must look in order to discover the true form of certain segments. What makes the task especially hard, though, is that this must all be done on the basis of incomplete learning data; waiting around to hear and memorize every form of every word would be impractical at best, and outright impossible in most cases. It seems safe to say that human learners must bring a number of different resources to the task: exquisite memories for large amounts of as yet unanalyzed data from the input, an ability to find and compare the relevant pairs and discover the patterns that are present in the data, a means of evaluating competing patterns to learn which are productive, an urge to generalize and project beyond the data, and a set of principles that govern how generalization proceeds.

To date, no model has been implemented that can take on the whole problem, even with idealized learning data consisting of complete paradigms. In this section, I outline a model of one piece of the larger problem: it compares different parts of the paradigm to figure out which is most revealing about the properties of words, and develops a grammar of morphological and phonological rules to project the rest of the paradigm. This model assumes that the learner has already performed tasks such as segmenting the speech stream into words, representing words in some type of phonemic representation, and arranging the words into sets of forms that are hypothesized to be morphologically related.[8] In addition, it assumes that the learner has already performed some preliminary phonological learning, in the form of discovering that certain sequences are non-occurring (surface illegal) in the language.[9]

In order to understand how the model calculates predictiveness and constructs grammars, it is useful to start with a schematic example.

[8] Some models that take on the task of word segmentation include Allen and Christiansen (1996), Brent and Cartwright (1996), Cairns, Shillcock, Chater, and Levy (1997), and Brent (1999). The task of finding pairs of words that are hypothesized to stand in a morphological relationship is less well understood, though see Baroni (2000) and Goldsmith (2001) for unsupervised approaches to morpheme discovery.

[9] A number of studies have shown that infants acquire some knowledge about sequence probabilities as early as eight months—well before they have significant knowledge of words or morphemes (Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk 1993; Friederici and Wessels 1993; Jusczyk, Luce, and Charles-Luce 1994). It is not unreasonable to suppose that this knowledge is brought to bear on the task of learning alternations between morphologically related words; see Hayes (2004) and Tesar and Prince (2007) for specific proposals along these lines.

7.3.1 Searching for contrastive information within paradigms

Learning to produce inflected forms of words would be a relatively easy task if all lexical items took the same sets of endings, there were no exceptional irregular forms, and phonology never acted to neutralize surface contrasts. In such a language, the learner would simply need to compare related forms of a few words in order to ascertain the suffixes. For example, faced with the paradigms in (7) (based on a simplified version of Middle High German), the learner could infer that the nominative singular suffix is null, the genitive singular suffix is *-es*, and the nominative plural suffix is *-e*.[10]

(7) Paradigms with no alternations

NOM SG	GEN SG	NOM PL	Gloss
stein	*steines*	*steine*	'stone'
sin	*sines*	*sine*	'sense, mind'
arm	*armes*	*arme*	'arm'
sal	*sales*	*sale*	'hall'
ʃreɪ	*ʃreɪes*	*ʃreɪ*	'cry, shout'

Actual languages can, of course, be more complicated: phonology may act to neutralize surface contrasts in some parts of the paradigm, words may fall into different inflectional classes, and there may be irregular exceptions that fail to follow any of the major patterns. Consider the following sets of forms (again, based loosely on Middle High German), in which some forms show voicing alternations in the stem-final consonant (8a), while others do not (8b).

(8) Phonological neutralization

a. Stems with voicing alternations

NOM SG	GEN SG	NOM PL	Gloss
tot	*todes*	*tode*	'death'
niːt	*niːdes*	*niːde*	'enmity'
helt	*heldes*	*helde*	'hero'
sant	*sandes*	*sande*	'sand'
tak	*tages*	*tage*	'day'
tswiːk	*tswiːges*	*tswiːge*	'branch'
diŋk	*diŋges*	*diŋge*	'thing'
vek	*veges*	*vege*	'way'
briəf	*briəves*	*briəve*	'letter'
kreɪs	*kreɪzes*	*kreɪze*	'circle'
lop	*lobes*	*lobe*	'praise'

b. Non-alternating stems

NOM SG	GEN SG	NOM PL	Gloss
muːt	*muːtes*	*muːte*	'courage'
ʃrit	*ʃrites*	*ʃrite*	'step'
knext	*knextes*	*knexte*	'servant'
geɪst	*geɪstes*	*geɪste*	'spirit'
nak	*nakes*	*nake*	'nape'
blik	*blikes*	*blike*	'glance'
druk	*drukes*	*druke*	'pressure'
lok	*lokes*	*loke*	'lock (hair)'
ʃif	*ʃifes*	*ʃife*	'ship'
slos	*sloses*	*slose*	'lock'
ʃimpf	*ʃimpfes*	*ʃimpfe*	'taunt'

[10] I leave aside the possibility that all nouns end in *-e* and that the nominative singular shows truncation, with *-s* and Ø as the genitive singular and nominative plural suffixes, respectively. Such parsing problems can be non-trivial, however. For example, given just the first set of forms (*jar*, *jares*, *jare*), the learner might be uncertain about whether the *r* is part of the stem or the suffix. The procedure described below operates under the assumption that if material is shared by all morphologically related forms, it is part of the stem—that is, *jar-Ø*, *jar-es*, *jar-e*.

The data in (8) suggest that the language has a phonological process (such as final devoicing) that neutralizes the contrast between voiced and voiceless obstruents word-finally. In order to discover this, the learner must make two types of comparisons. First, by comparing nominative *tot* with genitive *tod-es*, it can be seen that some sort of process is operating to create voicing alternations. Second, by comparing *tot* ~ *todes* with *mut* ~ *mutes*, one can infer that it is a devoicing process, and not a voicing one (or else we would expect genitive singular **mudes*). This is the basic logic behind many phonology problems: noticing that rules must operate in one direction rather than vice versa, because there is an unpredictable opposition which is neutralized in some forms but not in others. Furthermore, having discovered this, the learner also now knows that the nominative singular is not a reliable source of information regarding the voicing of stem-final obstruents; for this, one must look to a suffixed form.

In addition to phonological neutralizations, the learner must also contend with the possibility of inflectional classes, which are not always distinct in all forms. For example, alongside the forms in (7) and (8), the language may have words like those in (9), which differ by changing their vowel in the plural form, or by taking a different suffix, or both.

(9) Stems in different inflectional classes

NOM SG	GEN SG	NOM PL	Gloss
sak	*sakes*	*seke*	'sack'
korp	*korbes*	*kœrbe*	'basket'
rok	*rokes*	*rœke*	'coat'
liːp	*liːbes*	*liːber*	'body'
vort	*vortes*	*vort*	'word'
liət	*liədes*	*liəder*	'song'
lant	*landes*	*lender*	'land'

The problem of discovering inflectional classes is usually treated as a separate problem from that of finding phonological contrasts, but the considerations are the same: an unpredictable difference seen in one part of the paradigm (here, the plural) may be neutralized in another part of the paradigm (i.e., the singular), forcing learners to look to a particular part of the paradigm for the crucial distinguishing information.

The task, then, is to discover that a contrast seen in form A in the paradigm is neutralized in form B. One possible approach would be to establish correspondences between every segment in form A and form B, checking to make sure that the relations are always one-to-one bijections (x always maps to y). If a many-to-one relation is discovered (x maps to y in one word and to z in another), then we could infer that a contrast between y and z is neutralized to x in some environment. This is shown schematically in (10).

(10) Establishing correspondence between segments in related forms

a. One-to-one relation

$a_1 b_2 \mathbf{x}_3$ ~ $a_1 b_2 \mathbf{y}_3$-suffix (*stein* ~ *stein-es*)

$c_1 d_2 \mathbf{x}_3$ ~ $c_1 d_2 \mathbf{y}_3$-suffix (*sin* ~ *sin-es*)

b. One-to-many relation

$a_1b_2\mathbf{x}_3$	~	$a_1b_2\mathbf{y}_3$-suffix	(*tak* ~ *tag-es*)
$c_1d_2\mathbf{x}_3$	~	$c_1d_2\mathbf{z}_3$-suffix	(*druk* ~ *druk-es*)

This approach will discover neutralizations in any part of the paradigm, but it has one unappealing trait as a learning procedure: it presupposes that the learner is certain about the morphological analysis (*tag-es*, and not, say, *ta-ges*). The question, essentially, is whether the learner can be sure that *y* and *z* actually stand in a correspondence relationship with one another, or whether they belong to different morphemes that happen to put them in the same position in the word. Unless the learner can be sure that both segments belong to the same morpheme, it is impossible to infer that the opposition represents a phonological contrast that is neutralized elsewhere.

A different approach, which avoids this problem and mirrors the way that phonology students are often guided to the right solution, is to use an error-driven strategy: if adding a suffix to *abx* yields *aby*-suffix, then we incorrectly predict that adding the same suffix to *cdx* should yield **cdy*-suffix. For example, on the basis of forms like [stein] ~ [steine] and [nak] ~ [nake], we might expect the plural of [tak] to be *[take]. Since it is not, we know that something additional must be learned. There are numerous possibilities:

1. The difference between *y* and *z* is due to a doubly-conditioned phonological process, caused by *b* or *d* on the left and (at least some part of) the suffix on the right.
2. The difference between *y* and *z* is underlying, and a phonological process neutralizes them to *x* word-finally.
3. One of the outcomes is due to an irregular morphophonological process that affects only some words (in practice, this explanation is often handled in the same way as (2)).
4. The segments *y* and *z* actually belong to two different suffixes, and the words belong to distinct inflectional classes.

The plausibility of one hypothesis over another typically depends on considerations such as the number of words involved, the naturalness of the change, and so on. In the hypothetical example in (7)–(9), a phonological final devoicing analysis seems persuasive because of the number of words (and different segments) involved, the fact that the alternation cuts across different inflectional classes, and the naturalness of final devoicing as a phonological process. In order to conclude this with certainty, however, we need data from more than just a few words. Informally, analysts typically seem to assume that learners must entertain a number of hypotheses simultaneously, until there is enough data available to make one seem more likely than the others. In the next section, I outline a computationally implemented model that tries to do just this, bootstrapping preliminary information about morphology and phonology to evaluate competing hypotheses about how to account for apparent unpredictability.

7.3.2 Discovering contrastive information algorithmically

The premise of the current approach is that learners direct their attention to the part of the paradigm that provides as much information as possible about how to inflect words accurately—that is, avoiding errors like **helte* for *helde*, or **sake* for *seke* by learning that the former has an underlying /d/, and the latter is an umlauting (vowel-changing) stem. The procedure described in this section attempts to learn both phonological and morphological contrasts simultaneously, starting with a first-pass analysis of the morphological changes involved, attempting to learn some phonology, and using this to improve its morphological analysis. Finally, since it does not know where in the paradigm contrastive information may occur, it does this starting from every part of the paradigm, in order to determine which one yields the most accurate generalizations.

The input to the model is a set of paradigmatically related forms in phonetic transcription, such as the ones in (7)–(9) above. In order to permit generalizations about phonological environments, the model is also provided with a matrix of phonological feature values for the sounds of the language (that is, knowledge of phonological features is assumed to be known ahead of time, perhaps by being innate). In addition, the model is provided with knowledge about sequences that are surface illegal in the language, in the form of a list of non-occurring sequences. In the case of a language with no word-final voiced obstruents, this list would include sequences like [b#], [d#], [g#], etc.

As discussed above, a key observation in discovering neutralizations is the simple fact that neutralizations lead to ambiguity, and, thus, potential uncertainty. For instance, given a nominative singular form [mut], the learner is not certain whether the plural should be [mute], [mude], or even [myte], [myde], or some other form. Thus, if one were to construct a grammar that used the nominative singular as its input and tried to generate nominative plurals by rule, there would be some indeterminacy concerning voicing, vowel quality, and also the correct suffix to use. In such a case, the grammar might pick one of these outcomes as the regular outcome (for example, simply adding *-e* with no voicing or vowel change), but this would leave unexplained many irregular forms that took other patterns. Going from the plural to the singular, on the other hand, there is no ambiguity concerning final obstruent voicing: when the suffix is removed and the obstruent is put into final position, it must be devoiced.[11]

It is important to recognize that frequently the seriousness of an ambiguity can be mitigated by means of detailed rules that capture subgeneralizations about the patterns involved. In the sample data in (9), for example, we see that final obstruents

[11] Note that whichever direction is chosen, there may still be ambiguities concerning root vowel alternations—for example, singular [a] could correspond to plural [a] or [e], while plural [e] could correspond to singular [e] or [a]—though even here, the plural→singular mapping is less ambiguous (singular [u] may correspond to plural [u] or [y], but plural [y] almost always corresponds to singular [u]).

are never voiced when they follow a fricative (plural [knexte], [geɪste], but no forms like *[bexde], *[meɪsde]). At the same time, it happens that in this set of forms, [t] always voices after [n] ([sande], [lender] but no hypothetical *[bante], *[menter]). Stem-final [p] always voices ([lobe], but no *[rope]), while stem-final [pf] never does ([ʃimpfe], but no *[dimbve]). These small-scale generalizations (dubbed ISLANDS OF RELIABILITY in Albright 2002b) have the potential to recover a good deal of information about a contrast that has been neutralized. Furthermore, there is a growing body of experimental evidence showing that speakers are actually sensitive to such patterns (Zuraw 2000; Albright, Andrade, and Hayes 2001; Albright 2002b; Albright and Hayes 2003; Ernestus and Baayen 2003). Therefore, any attempt to estimate the seriousness of a neutralization must explore the possibility of predicting one's way out of it, by means of such small-scale generalizations.

The Minimal Generalization model of Albright and Hayes (2002) is a model of grammar induction that is designed to do precisely this. It takes pairs of morphologically related forms and compares them, attempting to find the most reliable generalizations it can about the mapping from one form to the other. It starts by taking each data pair and comparing the input and output, to determine what has changed and what is constant. The result is expressed as a word-specific rule, describing the mapping involved for just this one datum. For example, given the nominative singular and plural forms in (7)–(9), the Minimal Generalization algorithm would start by factoring each pair into a changing and non-changing portion, thereby determining that several changes seem to be involved. This is shown, for a subset of the data, in (11). At a first pass, the changing portion corresponds roughly to the affixes, and the constant portion can be considered the stem—though in cases where the voicing of the final obstruent is altered, we see that this is also included as part of the change in this initial parse.

(11) Factoring the input data into change and context

Input	Output	Restated as a word-specific rule
muːt	muːte	Ø → e / muːt __ #
ʃrit	ʃrite	Ø → e / ʃrit __ #
knext	knexte	Ø → e / knext __ #
geɪst	geɪste	Ø → e / geɪst __ #
arm	arme	Ø → e / arm __ #
ʃreɪ	ʃreɪe	Ø → e / ʃreɪ __ #
tot	tode	t → de / to __ #
niːt	niːde	t → de / niː __ #
helt	helde	t → de / hel __ #
tak	tage	k → ge / ta __ #
vort	vort	Ø → Ø / vort __ #

The next step is to generalize, by comparing word-specific rules that involve the same morphological change. For example, comparing the word-specific rules for *mut*~*mute*

and *ʃrit* ~ *ʃrite*, the model posits a new rule added *-e* after any stem that ends in a [t] preceded by a high vowel.

(12) Generalization over pairs of related rules

Change	Residue	Shared features	Shared segments	Change location	Shared segments
Ø → e	m	uː	t	__	#
Ø → e	ʃr	i	t	__	#
Ø → e	X	$\begin{bmatrix}+\text{syllabic}\\ +\text{high}\end{bmatrix}$	t	__	#

The precise generalization scheme is as follows: moving outward from the change location, any strictly identical segments are retained in the generalized rule, in the SHARED SEGMENTS term. Upon encountering a pair of mismatched segments, the model compares them to determine what feature values they have in common; these are retained as the SHARED FEATURES. Finally, if either of the rules under comparison has additional material left over, this is converted to a free variable (here, X). The search for shared material is carried out symmetrically on both the left and right sides. Here, the fact that the change is word-final is indicated by means of a shared word-edge symbol (#), but it could also be indicated simply by the lack of a free variable on the right side (meaning no additional material can be matched on this side).

The comparison in (12) happens to yield a rule that is scarcely more general than the word-specific rules that spawned it. When the process is iterated over the entire input set, however, comparison of heterogeneous input forms can lead to much broader generalizations, including even context-free rules. One pathway to context-free *-e* suffixation is shown in Figure 7.1.

The goal of generalization is not merely to discover which contexts a change applies in, but also to discover where it applies reliably. This is assessed by keeping track of a few simple statistics. For each rule, the model determines how many forms in the input

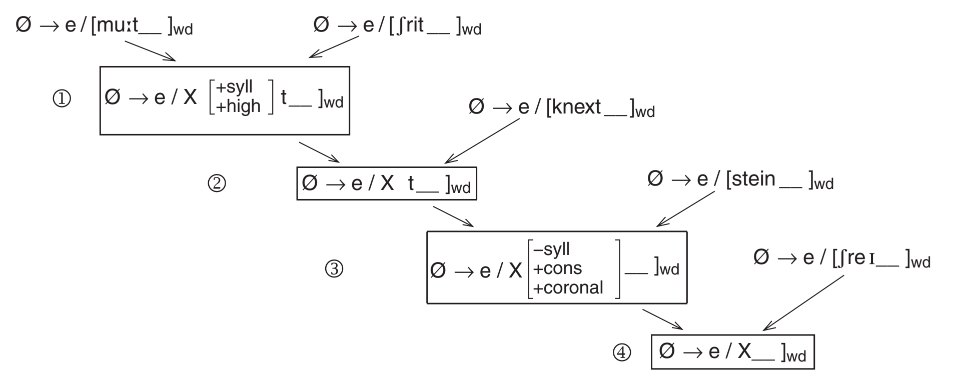


FIGURE 7.1. Iterative comparison yields broader generalizations

data meet the structural description of the generalization (data it tries to take on = its SCOPE), along with how many of those forms actually take the change required by the generalization (data it actually works for = its HITS). For example, consider the rule affixing [-e] after the sequence of a high vowel followed by [t]: Ø → e / [+syllabic, +high] t __ #. Among the forms in (7)–(9), four contain high vowels followed by [t] ([nit], [mut], [ʃrit], and [liət], provided we treat the diphthong [iə] as high), so the scope of the rule is 4. Only two of these words form their plural by simple [-e] suffixation, however ([mute], [ʃrite]), while [-e] suffixation incorrectly predicts *[nite], *[liəte] instead of correct [nide], [liəder]; so, the [-e] suffixation rule has two hits, and two exceptions. The RELIABILITY of a rule is the proportion of its hits to its scope: here, 2/4 = .5. By comparison, the rule changing [t] → [de] could potentially cover eleven of the forms in (7)–(9) ([tot], [nit], [helt], [sant], [mut], [ʃrit], [knext], [geɪst], [vort], [liət], [lant]), but only four of them actually have voicing alternations, so this rule would have a reliability of 4/11 = .36.

A model that values reliability above all else would try to find generalizations that have as few exceptions as possible, even if this meant carving the data up into a patchwork of small, independent generalizations to avoid admitting a few exceptions. In reality, we want to strike a balance between analyses that are accurate and those that are general enough to capture the data insightfully and can extend the patterns correctly to novel items. This is achieved in the Minimal Generalization model by adjusting reliability values downward using lower confidence limit statistics, to yield a CONFIDENCE score (Mikheev 1997). The result of this adjustment is that generalizations based on smaller amounts of data receive lower confidence scores: 2/2 = confidence of .52, 5/5 = confidence of .83, 20/20 = confidence of .96, and so on.[12] The relationship between the size of the generalization and the resulting confidence adjustment can be seen in Figure 7.2. By relying on confidence rather than on raw reliability, the model is able to favor broader generalizations (i.e., ones with more observations in their scope), even if they involve a few exceptions. As we will see in section 7.4.1, this adjustment also plays a crucial role when different amounts of data are available from different parts of the paradigm.

[12] The lower confidence limit of a reliability ratio is calculated as follows: first, the reliability (probability) ratio, which we may call $\hat{p}$, is adjusted to avoid zeros in the number or denominator, yielding an adjusted value $\hat{p}^*$: $\hat{p}^* = \frac{\text{hits} + 0.5}{\text{scope} + 1}$. This adjusted value is then used to calculate an estimate of the true variance of the sample:

$$\text{Estimate of variance} = \sqrt{\frac{\hat{p}^* \times (1 - \hat{p}^*)}{n}}$$

This value is then used to calculate the lower confidence limit (π_{Lower}), at a particular confidence value α:

$$\pi_{Lower} = \hat{p}^* - z_{(1-\alpha)/2} \times \sqrt{\frac{\hat{p}^* \times (1 - \hat{p}^*)}{n}}$$

The confidence value α ranges from $.5 < \alpha < 1$, and is a parameter of the model; the higher α is, the greater the penalty for smaller generalizations. In the simulations reported here, I will always assume an α of .75. The value z for a particular confidence level α is found by consulting a standard statistics look-up table.

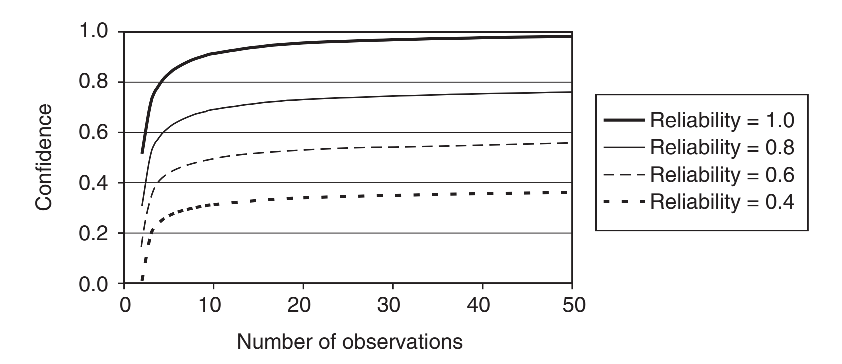


FIGURE 7.2. Relationship between amount of data and confidence limit adjustment

The process described thus far yields a rather uninsightful analysis of voicing alternations—namely, that words with voicing alternations constitute a separate inflectional class which take a different set of suffixes (singular *-t*, plural *-de*). The learner arrives at this analysis because the initial parse of the morphology occurs prior to any learning of phonological alternations, so there is no way of knowing that the [t] ~ [d] alternation could be explained on phonological grounds. During the course of assessing the reliability of generalized rules, however, there is an opportunity to improve on this analysis, in the following way: when the model discovers that a form meets the structural description of a rule but does not obey it, an error is generated, which can be inspected for phonotactic violations. If the incorrectly predicted form contains an illegal sequence, then there may be a phonological rule involved, and the model attempts to posit a rule that fixes the incorrect form, transforming it into the correct, observed one.

To take an example, when evaluating the morphological rule Ø → e / X [+syllabic,+high] t __ # discussed above, the model observes that the rule correctly generates the forms [mute] and [ʃrite] (two hits), but it incorrectly predicts the forms *[nite] and *[liəte] (for [nide] and [liəder], respectively). There are two possible reasons why the morphological rule generates the wrong outcome: either it doesn't apply to these words, or it does apply, but a phonological rule is needed to yield the correct surface output. Put more concretely, although the output *[nite] is incorrect, if the language had a process of intervocalic voicing, this would explain why the observed output is actually [nide]. The viability of an intervocalic voicing rule is tested by consulting the list of illegal sequences to see whether intervocalic [t] is known not to occur. In this case, we find that the hypothetical rule is not viable, since intervocalic [t] is fine in this language (in fact, it occurs in parallel forms like [ʃrite]). Thus, we correctly discover that if we take the nominative singular as our starting point, there is no more insightful analysis to be had; all we can say is that there is an irregular competing process that sometimes changes [t] to [d].

Let us now contrast this with an analysis using the nominative plural as an input. Here, the changes that we observe include removing a suffix (e.g., [e] → Ø), and removing a suffix with a concomitant voicing readjustment (e.g., [de] → [t], as in [tode] → [tot]). As above, iterative generalization discovers a range of possible contexts characterizing the two changes, and the model evaluates the reliability of all of these generalizations. Now the context-free rule for [e] → Ø makes the correct prediction for forms like [mute] and [ʃrite], but for plural [tode], [nide], and [tage], it incorrectly predicts singular *[tod], *[nid], and *[tag]. This time, the incorrect predictions could be fixed by a rule of final devoicing. When the erroneous predictions are compared against the list of illegal sequences, we find that final voiced obstruents are in fact illegal, and a final devoicing rule is viable. With a phonological devoicing rule in place, forms like [tot] and [nit] can be derived by the simpler [e] → Ø rule, and the reliability of this rule improves. Thus, by taking the plural as a starting point, the learner is able to come up with a unified analysis of the voiced and voiceless-final stems in (8).

We see from this example that when a form suffering from neutralizations is used as the input to the grammar, the resulting rules are less accurate and less reliable, since they have to make guesses about essentially unpredictable properties. This suggests a straightforward strategy for discovering which form in the paradigm exhibits the most contrasts: simply take each form in the paradigm and try learning grammars that derive the remaining forms from it. The slot in the paradigm that yields the grammars with greatest accuracy and most reliable rules is then chosen as the base of the paradigm, and the remaining forms are derived from it by means of morphological and phonological rules. Thus, the base form is chosen in order to maximize confidence in the remainder of the paradigm.

As noted in section 7.2.2, the confidence maximization approach has the potential to explain why analogical change sometimes takes the typologically unusual step of rebuilding more frequent, less marked forms. In the current example, based on Middle High German, we see that final devoicing made the nominative singular a relatively unpredictive form, while the number of different plural classes would have encouraged selecting a plural form as a base. This prediction seems to be borne out for real Middle High German: both Modern German and Yiddish show leveling of vowel length from the plural form (Paul, Wiehl, and Grosse 1989: §23), while Yiddish and some Bavarian dialects show the additional leveling of final obstruent voicing, discussed in section 7.2 above.

7.3.3 The single surface base hypothesis

In the example in the previous section, all phonological and morphological contrasts were aligned so they were most clearly visible in the same part of the paradigm (the plural). Frequently this is not the case, however. In fact, different parts of the paradigm often maintain different information, since phonological and morphological neutralizations can theoretically target any slot within the paradigm. For this reason,

it is generally assumed that learners are able to compare multiple forms of inflected words to arrive at their lexical representation (Kenstowicz and Kisseberth 1977).

In Albright (2002a), a more restrictive model of acquisition is proposed: when no single part of the paradigm maintains all contrasts, the learner is forced to choose the single form that is generally most predictive, even if this means losing information about certain contrasts. This constraint can be called the SINGLE SURFACE BASE HYPOTHESIS, since it requires that all paradigms of all words be organized around the same base form.

To see how the single surface base constraint works, let us consider some additional data from the history of German. At some point in Old High German or early Middle High German, the phoneme [h] (from older [χ] or [x]) was lost intervocalically (Braune and Mitzka 1963: §152b; Paul, Wiehl, and Grosse 1989: §111, §142). This created paradigmatic alternations, still seen in Modern German *hoch* [hoːx], *höher* [hœːɐ], *am höchsten* [hœːçstən] 'high/higher/highest'. It also created alternations in noun paradigms, since [h] deleted in forms with vowel-initial suffixes, such as the plural:

(13) Stems ending in (older) [h] ~ [x]

NOM SG	NOM PL	Gloss
ʃux	*ʃuː(w)e*	'shoe'
rex	*reː(j)e*	'deer'
flox	*flœːe*	'flea'

This change created a neutralization in the plural with nouns that did not have historical /h/.

(14) Stems without historical /h/

NOM SG	NOM PL	Gloss
kuː	*kyːe*	'cow'
weː	*weː(j)e*	'woe'

Furthermore, an independent change of older [k] > [x] in Old High German also worked to create a neutralization in the singular, with non-alternating [x] nouns.

(15) Stems without historical /k/

NOM SG	NOM PL	Gloss
bux	*byxer*	'book'
bax	*bexe*	'stream'
pex	*pexe*	'pitch'
kox	*kœxe*	'cook'

Such nouns went against the overall trend for contrasts to be preserved more faithfully in the plural; in fact, in each part of the paradigm, there is one contrast and one neutralization (the NEUTRAST configuration; Kager, in press). In order to learn whether a word had stem-final /h/ or not, learners would have had to compare also the nominative singular for this set of words.

For a learner operating under the single surface base restriction, however, such a strategy is not possible. Since the nominative plural form is most informative about the

majority of other contrasts in the language, it must serve as the base for this subset of words as well.[13] As a consequence, this class of words must be stored without the [x] or /h/. Furthermore, since Middle High German had so few words with historical /h/, the rule "restoring" it in the nominative singular (e → x / [+syllabic] __ #, producing derivations like [ʃuːe] → [ʃux]) would have had extremely low confidence. Thus, under this restriction, the grammar of MHG could not possibly have generated forms like [ʃux] or [flox] productively with high confidence rules. The only way to produce such forms would have been to memorize them as irregular exceptions, in order to block the grammatically expected forms [ʃuː], [reː], [floː].

The single surface base restriction seems drastic, but it makes the right prediction. If we assume that errors (by children or by adults) are overwhelmingly overregularizations (that is, replacement of irregular forms by grammatically expected regular forms), then we predict that older /h/ words should have lost the [x] in the nominative, as innovative regularizations ([floː]) gradually replaced the older irregulars ([flox]). We do not predict the converse changes of importing [x] to the plural ([flox] ∼ *[flœxe], on analogy with [kox] ∼ [kœxe]) or deleting the [x] in historical /k/ words ([kox] ∼ [kœe], on analogy with [flox] ∼ [flœe]). Under this model, the failure of [flox]-type words to interact with [kox]-type words is due to the simple fact that historical /h/ and /k/ words remained distinct in the plural ([flœe] vs. [kœxe]). And in fact, this prediction is borne out: words like [ʃux], [rex], and [flox] lost their [x] by analogy (Molz 1906: 294; Paul, Wiehl, and Grosse 1989: §25c, 44), and are pronounced [ʃuː], [reː], and [floː] in Modern German, while historically vowel-final and *k*-final words remained unchanged.

This example shows how the current approach can make very specific predictions about particular instances of analogical change. It predicts not only which form in the paradigm will be affected (non-basic forms, which are open to rebuilding if they cannot be generated correctly by grammar), but also which direction the change will go in (regularization to the lexically dominant pattern). By using a synchronic model of paradigm acquisition to predict asymmetries in possible errors, we are able to achieve a more constrained and explanatory theory of the direction of analogy.

This model has been shown to work in several other unusual cases of analogy, as well. In Albright (2002a), I showed that it made the right predictions for three typologically unusual paradigmatic changes. The first was a case of across-the-board leveling to the first-person singular in Yiddish verbs, in violation of the tendency to level to the third-person singular. In this case, the advantage of the first-person singular appears to be due to the devoicing or even total loss of stem-final obstruents which occurs in the second- and third-person singular and the second-person plural, together with the fact that the first-person singular maintains a contrast in stem-final schwas which is sometimes difficult to recover from the first- or third-person plural or infinitive. Since

[13] This discussion is predicated on the (almost certainly true) assumption that words ending in obstruents outnumbered words ending in historical /h/ in Middle High German.

the first person singular is the only form that maintains all of these contrasts, it is the most predictive, and is (correctly) chosen to serve as the base. The second unusual change involved the elimination of [s] ~ [r] alternations in Latin (the famous *honor* analogy), in which the nominative singular form of noun paradigms was rebuilt on the basis of an oblique form. As with the MHG case discussed here, the preference for a suffixed form in Latin seems to be due to phonological processes that affected word-final obstruents. The details of the change, which affected only polysyllabic non-neuter nouns, are also correctly predicted by a model that uses probabilistic rules to capture lexical tendencies in different contexts. The final case involved an analogical change in Lakhota verbs which appears to have been based on the second-person singular. In this case, the neutralizations involved were both more complex and more symmetrical, but the advantage of the second-person singular seems to have come from the fact that it maintained the contrast between two large classes of words which were neutralized elsewhere in the paradigm. For details on all of these changes, the reader is referred to Albright (2002a).

The upshot of this section is that the proposed model makes advances in explaining the language particulars of analogical change. What remains to be shown, however, is whether it has anything to say about universal tendencies.

7.4 TYPOLOGICAL TENDENCIES: EXPLORING THE PARAMETER SPACE OF THE MODEL

The model laid out in the previous section makes a strong claim about base forms in paradigms. It posits that bases play an integral role in the synchronic organization of grammar, and that they are chosen in order to facilitate, or optimize, the resulting grammar. A base form is considered optimal in this system if it contains enough information to reliably predict the remaining forms in the paradigm, by preserving contrasts and lacking neutralizations. This procedure for selecting base forms seems rational as a theory of how synchronic grammars are organized, and also makes the correct predictions for individual cases which are typologically unusual. In this section, I show that this procedure also makes the correct typological predictions.

There are two distinct issues that must be addressed in assessing the typological predictions of the model. The first is the issue of empirical coverage: are there attested analogies which run counter to the hypothesis that base forms are always the most informative form? The second issue is one of relative frequency: in principle, contrasts could be maintained anywhere in the paradigm (first-person singular, second-person singular, third-person singular, etc.), and in many cases, contrasts are maintained equally well by multiple forms. Why is there a strong tendency for analogy to be based on the most frequent or least marked forms?

I consider each of these questions in turn.

7.4.1 Leveling to uninformative base forms

A common criticism of proposed principles of analogy is that there always seem to be exceptions: even if analogy usually extends the most frequent, the least marked, or the unsuffixed form, occasionally it goes in the opposite direction, extending a less frequent, more marked, suffixed form. Under a tendency-based approach, such exceptions are not necessarily a problem, since the goal is to explain only what is *likely*, not what is *possible*. The current model makes a stronger claim, however, that the direction of analogy should be predictable in all cases. The question immediately arises, therefore, whether there are exceptions to the informativeness-based account of analogy, just as there are exceptions to every other proposed tendency.

There are in fact a number of well-known examples of analogies based on pivot forms that involve massive neutralizations. One case that is often cited comes from Maori (Hohepa 1967; Hale 1973a; Kiparsky 1978; Hock 1991: 200–202; Blevins 1994; Barr 1994: 468–477; Kibre 1998). In Maori, passives were historically formed by adding a vowel-initial suffix (generally *-ia* or *-a*) to the verb stem: *awhit* → *awhit-ia* 'embrace', *hopuk* → *hopukia* 'catch', and so on. Subsequently, word-final stops were deleted (*awhit* > *awhi*), creating alternations within verb paradigms. (For more detailed discussion of Maori and comparative data from related languages, see Blevins, this volume.)

(16) Unpredictable consonants in Maori passive

Verb	Passive	Gloss
awhi	*awhi***t***ia*	'embrace'
hopu	*hopu***k***ia*	'catch'
aru	*aru***m***ia*	'follow'
waha	*waha***ŋ***ia*	'carry on back'
mau	*mau***r***ia*	'carry'
wero	*wero***h***ia*	'stab'
hoka	*hokaia*	'run out'
patu	*patua*	'strike, kill'

The Maori case is interesting because the passive suffix has apparently been reanalyzed as a set of competing consonant-initial suffixes (*-tia*, *-kia*, *-mia*, *-ŋia*, *-ria*, *-hia*, etc.), or combinations of "thematic consonants" + *(i)a* suffix. As Blevins and others point out, a likely mechanism for this reanalysis is rule inversion, in which the unsuffixed form of the verb is taken as basic, and the speakers are then forced to learn and remember which consonantal suffix each verb takes. Unsurprisingly, this has led to significant analogical restructuring, with the *-tia* and *-a* suffixes gradually replacing the remaining allomorphs in most contexts: for example, newer *wahatia*, *wahaa* alongside older *wahaŋia*. The mystery for the current approach is why the unsuffixed form would ever be considered basic, even though it lacks information about unpredictable final consonants.

A similar change is underway in present-day Korean, in which noun paradigms are being rebuilt on the basis of unsuffixed (isolation) forms, even though these forms suffer from drastic coda neutralizations. In Korean, all obstruents are realized as

unreleased stops in coda position; thus, for example, underlying /nat/ 'grain, kernel', /nat^{h}/ 'piece', /nas/ 'sickle', /nac/[14] 'daytime', and /nac^{h}/ 'face' are all pronounced [nat˺]. In suffixed forms, however, stem-final consonants are intervocalic, and underlying contrasts are generally preserved in conservative speech and in standard Korean orthography (e.g., accusative [nad-ɨl], [nat^{h}-ɨl], [nas-ɨl], [naj-ɨl], [nac^{h}-ɨl]). The result is that noun paradigms frequently contain alternations.

(17) Alternations in Korean nouns (conservative pronunciation)

Unmarked	Accusative	Gloss
nat˺	*nadɨl*	'grain'
nat˺	*nat^{h}ɨl*	'piece'
nat˺	*nasɨl*	'sickle'
nat˺	*najɨl*	'daytime'
nat˺	*nac^{h}ɨl*	'face'

These underlying contrasts are gradually being lost in present-day Korean, but interestingly, the change has gone in the direction of replacing all stem-final coronals with /s/ or /c^{h}/: older [nat^{h}-ɨl] ⇒ newer [nas-ɨl], [nac^{h}-ɨl] (K. Ko 1989 [cited by Kang 2002]; Hayes 1995; Kenstowicz 1996; H. Kim 2001: 104; Kang 2002, 2003; Davis and H. Kang 2003; H. Ko 2006; Lee, n.d.; Sohn, n.d.; and many others).

(18) Analogical innovations in suffixed forms

Unmarked	Conservative accusative form	Present-day variation	Gloss
k'ot˺	*k'oc^{h}-ɨl*	*k'oc^{h}-ɨl, k'os-ɨl*	'flower'
pat˺	*pat^{h}-ɨl*	*pat^{h}-ɨl, pac^{h}-ɨl, pas-ɨl*	'field'
cət˺	*cəj-ɨl*	*cəj-ɨl, cəs-ɨl*	'milk'

It appears that the [t˺] ~ [s] and [t˺] ~ [c^{h}] alternations are being analogically extended, and that the pivot, or base of the change is the unmarked form ending in [t˺].[15] Similar changes are also taking place at other points of articulation, with [p] and [k] gradually replacing [p^{h}] and [k^{h}]. Like the Maori example, this analogy appears to be based on a far *less* informative base form.

What are we to make of such exceptions? One possibility would be to concede that informativeness, like frequency, markedness, and so on, is just one of many factors

[14] The precise realization of the obstruents transcribed as /c/, /c^{h}/, and /c'/ is subject to a fair amount of individual and dialect variation. They are considered by some to be alveolar affricates (Cho 1967: 45; Kim 1999), by others to be postalveolar or palatal affricates (Kim-Renaud 1974; Ahn 1998: 55), and by yet others to be palatal stops (e.g., Sohn 1999: 153); see Martin (1992: 28-29) for discussion. I follow majority opinion in calling them [−anterior] affricates, but nothing in the discussion that follows depends on their exact featural representation. I will (non-standardly) use [j] to indicate the intervocalically voiced counterpart of lax [c].

[15] An alternate possibility, advocated by Kim (2001), is that the change from [t, t^{h}, c, c^{h}] to [s] is simply a phonetic sound change. Kang (2003) provides detailed empirical evidence showing that although the change to [s] is most advanced in environments where it is phonetically natural, it is also taking place in other, unnatural environments. (See also Comrie 1979 and Garrett, this volume, on the related issue of natural alternations resisting analogical change.) Furthermore, a purely phonetic account could not explain why the change to [s] is restricted to nouns.

that determine the direction analogy, and that it, too, has exceptions. I believe this conclusion is premature, however, and that even these apparent counter-examples can be handled within an informativeness-based approach. In subsection (i) I will show that when the informativeness of the various Korean noun forms is calculated, it turns out that unmarked forms are, surprisingly, more informative than suffixed forms. The reason is twofold: first, the coda neutralizations shown in (17) actually cause very little ambiguity in practice, due to statistical asymmetries in the lexicon of Korean. Second, unmarked forms are overwhelmingly more frequent than suffixed forms, so learners receive much more learning data about them. Once the lexical statistics of Korean and the relative availability of forms are taken into account, it emerges that the isolation form is actually the most reliable base form for Korean.

(i) Neutralizations affecting Korean noun paradigms

Korean nouns are marked for case, as shown in (19). The case markers (referred to variously as suffixes or as particles in the literature) correspond roughly to functions such as nominative, accusative, dative, etc., though their distribution and syntax differ considerably from their equivalents in Indo-European languages (Sohn 1999: 231f., 326–350). In addition, case marking is optional in many syntactic contexts, and nouns may occur in unmarked (bare) forms. Finally, many cases have separate markers for consonant- and vowel-final nouns, such as nominative *-ka* (after consonants) vs. *-i* (after vowels). Partial paradigms for representative consonant and vowel final nouns are given in (19).

(19) Korean case marking for consonant- and vowel-final nouns (partial list)

	/k^{h}oŋ/ 'bean'	/k^{h}o/ 'nose'
Unmarked	*k^{h}oŋ*	*k^{h}o*
Nominative	*k^{h}oŋ-i*	*k^{h}o-ka*
Accusative	*k^{h}oŋ-ɨl*	*k^{h}o-ɾɨl*
Dative/Locative	*k^{h}oŋ-ey*	*k^{h}o-ey*
Genitive	*k^{h}oŋ-ɨy*	*k^{h}o-ɨy*
Topic	*k^{h}oŋ-ɨn*	*k^{h}o-nɨn*

Noun inflection is complicated by the fact that Korean has a number of phonological processes that give rise to alternations within the paradigm. Perhaps the most salient of these is the neutralization of all obstruents to unreleased stops in coda position. The obstruent inventory of Korean, which is shown in Table 7.1, contains several different manners (stops, affricates, and fricatives), as well as three different laryngeal settings (plain unaspirated [C], aspirated [C^h], and tense, or fortis [C']). Among labials and velars, this results in a three-way contrast (/p, p^h, p'/, /k, k^h, k'/), while among coronals, there are eight different phonemes: /t, t^h, t', t͡ʃ, t͡ʃ^h, t͡ʃ', s, s'/.

Word-initially, all of the laryngeal and manner distinctions are contrastive. In coda position, however, only unreleased stops are allowed: [p˺, t˺, k˺]. As as result, when a

TABLE 7.1. Korean obstruent inventory

		Labial	Coronal	Velar
Stops	Unaspirated (lenis)	p	t	k
	Aspirated	p^h	t^h	k^h
	Tense (fortis)	p'	t'	k'
Affricates	Unaspirated (lenis)		t͡ʃ	
	Aspirated		t͡ʃ^h	
	Tense (fortis)		t͡ʃ'	
Fricatives	Unaspirated (lenis)		s	
	Tense (fortis)		s'	

noun is unmarked for case, the final consonant of the stem is subject to neutralization. For example, the three-way contrast between /k/, /k^h/, and /k'/, which is maintained in suffixed forms (at least in conservative speech), is neutralized to [k˺] in unsuffixed forms (20a). The same is true in principle for the labial stops /p/, /p^h/, and /p'/, although in practice there seem to be no noun stems ending in /p'/ (20b).

(20) a. Neutralization of /k/, /k^h/, and /k'/ in unmarked forms

UR	Unmarked	Nominative	Accusative	Gloss
/c^{h}uək/	*c^{h}uək˺*	*c^{h}uəg-i*	*c^{h}uəg-ɨl*	'reminiscence'
/puək^h/	*puək˺*	*puək^h-i*	*puək^h-ɨl*	'kitchen'
/pak'/	*pak˺*	*pak'-i*	*pak'-ɨl*	'outside'

b. Neutralization of /p/ and /p^h/ in unmarked forms

UR	Unmarked	Nominative	Accusative	Gloss
/ip/	*ip˺*	*ib-i*	*ib-ɨl*	'mouth'
/ip^h/	*ip˺*	*ip^h-i*	*ip^h-ɨl*	'leaf'

Among coronal obstruents, coda neutralization is even more severe, as the [nat˺] example in (17) above shows. In this case, however, the full set of underlying contrasts is not visible in the nominative either, due to an additional process of palatalization before the high front vowel [i]. This palatalization changes /t/ and /t^h/ to [c] and [c^h], respectively, before [i] in derived environments.

(21) Neutralization of coronals in unmarked and nominative forms

UR	Unmarked	Nominative	Accusative	Gloss
/nat/	*nat˺*	*naj-i*	*nad-ɨl*	'grains'
/nac/	*nat˺*	*naj-i*	*naj-ɨl*	'daytime'
/nat^h/	*nat˺*	*nac^h-i*	*nat^h-ɨl*	'piece'
/nac^h/	*nat˺*	*nac^h-i*	*nac^h-ɨl*	'face'
/nas/	*nat˺*	*naʃ-i*	*nas-ɨl*	'sickle'

From the distribution of contrasts in (21), we see that stem-final [t, t^h, c, c^h, s][16] are completely distinct only when the suffix begins with a vowel other than [i], such as in

[16] There are no noun stems ending in the tense coronal obstruents [t', c', s'].

the accusative. Therefore, as far as final obstruents are concerned, it appears that the accusative would be the most reliable source of information about the properties of a Korean noun, even though it is suffixed, less frequent (see below), and more marked.

There are several reasons why this picture of Korean noun paradigms is incomplete, however. First, the coda neutralizations in (20)–(21) are not nearly as serious as they appear. In principle, coda [t˺] could correspond to any of [t, t^h, t’, t͡ʃ, t͡ʃ^h, t͡ʃ’, s, s’]. In practice, however, most of these are rare or unattested word-finally. In (22), corpus counts are given from the 43,932 nouns in the Sejong project corpus[17] (Kim and Kang 2000). These counts may be taken to represent an older stage of Korean, since nouns in the corpus are listed in standard Korean orthography, which is relatively conservative in representing final obstruents.

(22) Distribution of final obstruents

a. Labials		b. Coronals		c. Velars	
p	1360	t	1	k	5994
p^h	64	t^h	113	k^h	18
p’	0	t’	0	k’	6
		c	17		
		c^h	160		
		c’	0		
		s	375		
		s’	0		

These counts reveal several important facts. First, obstruent-final nouns are somewhat underrepresented, accounting for a total of only 8,108, or roughly 18 percent of the nouns in the corpus. This can be compared with 17,344 vowel-final stems (39 percent), and 18,477 sonorant-final stems (42 percent). Within the obstruents, there are drastic asymmetries between different places of articulation. First, coronals are relatively underrepresented in comparison to the other places of articulation.[18] Furthermore, among labials and velars, plain (unaspirated) stops are most common stem-finally, particularly among velars. By contrast, among coronals, [t] is virtually unattested, while almost 80 percent of the stems end in /s/ or /c^h/.

These facts conspire to make final segments far more predictable than the schematic examples in (20)–(21) would suggest. For the 80 percent of the vocabulary that ends in vowels or sonorant consonants, the final segment surfaces faithfully in all parts of the paradigm, including the unmarked form. For labial and velar obstruents (17 percent of the vocabulary), [p˺]/[k˺] in the unmarked form corresponds to plain [p]/[k] in suffixed forms 99 percent of the time. The only nouns for which the unsuffixed form is truly

[17] http://sejong.or.kr

[18] The magnitude of this difference is somewhat smaller if Sino-Korean items are excluded, since Sino-Korean final [p] and [k] were borrowed faithfully, but final [t] was adapted as [l]. There is no reason to exclude these items from discussion (since the changes discussed here happened well after the influx of Sino-Korean borrowings), but even if they are excluded, it still appears to be the case that coronals are surprisingly underattested.

uninformative are those that end in [t˺] (just 1.5 percent of the vocabulary), and even here, chances are high that the stem ends in [s] or [c^h] in suffixed forms.

In order to confirm the fact that final segments are, by and large, predictable in Korean, I carried out a simulation using the minimal generalization learner to learn Korean noun paradigms. Nouns from the Sejong corpus were romanized using the "hcode" Hangul Code Conversion software (Lee 1994), and then converted to a broad phonetic transcription using the set of phonological rules outlined by Yoon, Beckman, and Brew (2002) for use in Korean text-to-speech applications. Finally, input files were created containing the two hundred most frequent nouns, in their unsuffixed, nominative (-i/-ka), and accusative (-ɨl/-ɾɨl) forms. The minimal generalization learner was then trained on the task of predicting accusative forms on the basis of either unmarked or nominative forms.

Unsurprisingly, the results show that it is easier to predict the accusative form using the nominative (suffixed) form than the unsuffixed form, because there are fewer neutralizations involved. It is worth noting, however, that the difference is extremely small, owing to the small number of words that are actually affected by these neutralizations, and the possibility of guessing strategically based on the neighboring segmental context. The numbers in (23) confirm the claim that suffixed forms of Korean nouns are extremely predictable no matter what form is used as the base: whichever direction is chosen, it is possible to learn a grammar that predicts the correct outcome over 97 percent of the time, and employs rules that have very high confidence values attached to them (>.97 out of 1). Virtually identical results were obtained using input files of different sizes (the 100, 500, and 1,000 most frequent nouns).

(23) Relative informativeness of unmarked vs. nominative forms

	Accuracy of grammar	Mean confidence
Unmarked → Accusative	97.5%	.971
Nominative → Accusative	98.6%	.986

Finally, it is necessary to consider the informativeness of the accusative form as a potential base, since it preserves all obstruent contrasts, and thus appears to be even more informative than the unmarked or nominative forms. As it turns out, this is not true when we consider the lexicon as a whole, rather than just obstruent-final nouns. The reason is that the accusative marker has two shapes: *-ɨl* after consonants, and *-ɾɨl* after vowels. The two liquids [l] and [ɾ] are in allophonic distribution in Korean: [l] occurs in coda position, and [ɾ] occurs intervocalically. Thus, when a noun ends in the consonant /l/, it takes the *-ɨl* allomorph of the accusative marker, and the stem-final /l/ becomes [ɾ]: /il-ACC/ 'work' → [iɾɨl]. However, this is the same result that one would get from a vowel-final stem followed by the *-ɾɨl* allomorph: /i-ACC/ 'teeth' → [iɾɨl]. As a consequence, accusative forms ending in [ɾɨl] are potentially ambiguous: they could be parsed as an /l/-final stem with the *-ɨl* allomorph, or a vowel-final

stem with the *-ɾɨl* allomorph.[19] Furthermore, this situation could arise quite often. Counts from the Sejong corpus show that 17,344 out of 43,932 nouns (39 percent) end in vowels, while 3,472 (8 percent) end in /l/. This means that up to 47 percent of accusative forms end in *-ɾɨl*, and are thus ambiguous with respect to how the remainder of the paradigm should be formed.

The relative unreliability of the accusative as a base form was confirmed by carrying out learning simulations in the accusative→unmarked and accusative→nominative direction, again on the two hundred most frequent words in the Sejong corpus. As the table in (24) shows, using the accusative to project the unmarked and nominative forms does lead to more mistakes and more uncertainty than the other way round, because of *-ɨl*/*-ɾɨl* ambiguities. It might also be noted that in spite of these errors, overall performance is nonetheless quite good. This is an example of what Hayes (1999) calls MULTIPLE PREDICTABILITY: all of the forms in the paradigm are mutually predictable to a greater extent than is logically necessary.

(24) Relative informativeness of unmarked vs. nominative forms

	Accuracy of grammar	Mean confidence
Accusative → Unmarked	93.0%	.929
Accusative → Nominative	93.5%	.932

The overall informativeness of the unmarked, nominative, and accusative forms is summarized in Figure 7.3, which shows the average confidence with which each form may be used to predict the remaining two forms. The graph serves to reiterate the point that once we move beyond just obstruent-final nouns and look at the language as a whole, the unmarked and nominative forms are actually more informative than the accusative, even though they appear at first glance to suffer from more serious neutralizations. This finding is significant for two reasons: first, it is a vivid demonstration of how informativeness, in the sense that I am using it here, is difficult to intuit on the basis of schematic data, and can be calculated for certain only by running simulations on representative learning data. More important, this result brings us one step closer to understanding the Korean change, since it reveals that although the unmarked form is not the most informative form, it is extremely close (a 1 percent difference).

One fact that we have not yet considered is the relative token frequency of the various noun forms in Korean. Token frequency does not play any direct role in the model, as it was described in section 7.3. However, token frequency can play an important indirect role in determining what data a learner is likely to encounter. In the following section, I will show that differences in token frequency make the unmarked

[19] Likewise [pʰaɾɨl] from /pʰa/ 'green onion' or /pʰal/ 'arm', [naɾɨl] from /na/ 'I' or /nal/ 'day', [saɾɨl] from /sa/ 'four' or /sal/ 'flesh', [maɾɨl] from /ma/ 'the south' or /mal/ 'horse', and many others. It should be noted that Korean orthography does differentiate stem-final vs. suffix-initial /l/, but this is a purely morphemic (and not phonetic) distinction.

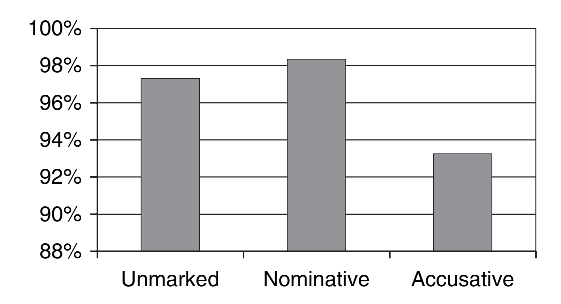


FIGURE 7.3. Relative informativeness of Korean noun forms, as measured by mean confidence in projecting the remainder of the paradigm

form a more reliable base form in Korean, in spite of the fact that it is slightly less informative.

(ii) Availability of forms: the influence of token frequency

The simulations in the previous section assume that all words are equally available to the learner in all forms (unmarked, nominative, accusative, etc.). In real life, however, learners plainly do not have equal access to all forms. For more frequent parts of the paradigm, lots of input data will be available, including forms of both common and rare lexical items. For less frequent parts of the paradigm, however, less input data is available, and on average only the more common words will have been encountered.

In Korean, the frequency difference between the various inflected forms is striking. Case marking is often omitted, particularly on accusative nouns. The relative token frequency of unmarked, nominative-marked, and accusative-marked nouns taken from counts of child-directed speech is given in (25) (Lee 1999). (Counts from non-child-directed speech show a similar, but slightly less drastic frequency difference.)

(25) Relative frequency of Korean nouns forms in child-directed speech (Lee 1999)

Form	Approximate % of all tokens
Unsuffixed	75
Nominative	20
Accusative	5

The relative lack of data for forms like the accusative has an indirect impact on its reliability as a base form. Recall from Figure 7.2 above that in the rule evaluation scheme employed by the minimal generalization learner, there is a statistical confidence adjustment that rewards rules based on more data. The effect of this adjustment is to penalize rules that have been tested against just a few forms. Its main purpose is to reward patterns that are well-instantiated, by giving them slightly higher confidence, and hence a greater productivity. A side effect of this adjustment, however, is that it rewards parts of the paradigm that are well instantiated, since their grammars are based on more data, and thus receive overall less downward adjustment. Might this

penalty be sufficient to tip the scales in favor of selecting a slightly less informative, but more frequent form as a base?

In order to test this, another simulation was performed, this time taking token frequency into account. Starting once again with the 43,932 nouns in the Sejong corpus, a simulated "Parental Locutionary Device" was constructed, which randomly produced inflected noun forms, chosen in Monte Carlo fashion according to both their lexical frequency and the relative frequency of the inflection. The probability of choosing a particular inflected form was calculated as in (26).

(26) $P(\text{inflected form}) = P(\text{lemma}) \times P(\text{inflection})$

The simulated parent produces common words more often than uncommon words, and frequent parts of the paradigm more often than infrequent parts of the paradigm. When asked to produce 1,000 tokens of Korean nouns, on average it produces 750 unmarked tokens, 200 nominative tokens, and 50 accusative tokens, respecting the relative frequencies in (25). Many lexical items happen to be produced in both unsuffixed and nominative forms, since these are both relatively frequent parts of the paradigm. The probability of producing the same word in both the unsuffixed and accusative forms, on the other hand, is much smaller (since accusative forms are simply not produced very often), and the probability of producing the same word in both the nominative and accusative is even smaller. Of course, the simulated parent also sometimes produces duplicate tokens, repeating the same form of the same word two or more times in the space of the same 1,000-word text. Duplicate tokens were not counted as separate learning data (i.e., the text was converted to types for morphological learning.)

Since chance can produce radically different results on different occasions, the simulated parent was used to produce ten different texts of 1,000 tokens each. These sets of forms were then used as learning data for the minimal generalization learner, in ten "simulated childhoods". Recall that the question of interest here is whether the relative frequency of unmarked forms over nominative ones is enough to render it a more reliable base form. The results of these simulations, given in (27), show that the frequency difference is more than sufficient to produce the desired result: the nominative form receives significantly lower scores, even though it is slightly more revealing about underlying phonemic contrasts.

(27) Average scores for unmarked and nominative forms, taking token frequency into account

	Mean confidence	Average winning margin
Unmarked → Accusative	.795	.812
Nominative → Accusative	.461	.475

The upshot is that once token frequency is taken into account to provide more realistic learning data, the most reliable base form in Korean is in fact the unmarked form. This result contradicts initial impressions of Korean noun paradigms based solely on

schematic data, and is due to three separate factors. The first is the relative rarity of stem-final coronal obstruents in the lexicon, meaning that the phonological rule of coda neutralization affects far fewer words than one might expect. Second, forms without coda neutralization suffer from their own independent neutralizations (such as palatalization, or the *-ɨl*/*-ɾɨl* ambiguity). Finally, the extreme frequency difference between unmarked and marked forms makes data about suffixed forms quite sparse for the learner.

The Korean case is illustrative in two respects: first, it shows that informativeness about phonological and morphological properties is complex to evaluate, and typically requires quantitative assessment of competing processes. Second, it shows that the assessment of reliability is based on more than just simple informativeness; a good base form must be not only informative about lexical contrasts, but also adequately available. I conjecture that this example is typical, and that when analogies appear to employ uninformative base forms, those forms are more informative than it seems, and are often substantially more frequent than the rest of the paradigm.[20]

The claim that there is a bias towards using more frequent forms as bases is certainly not a new one. It echoes suggestions by Mańczak (1958), Bybee (1985, 2001), and others, who point out that frequent forms are often the pivots for analogical change. The current model differs from previous proposals in the explanation for this bias, however. According to Bybee, token frequency plays a direct role in determining the organization of paradigms: as certain forms of the paradigm are heard and used more often, their lexical strength increases, and this makes them more influential and more basic (Bybee 1985: 117). By contrast, in the current model the grammar does not care about token frequency *per se*, but only about confidence of mappings. The relative token frequency of forms is not encoded anywhere in the grammar, nor does the model have any explicit bias to select frequent paradigm members as bases. The learner simply wants to select a base form that preserves as many contrasts as possible, but occasionally, the ideal form is so infrequent that it is impossible to be confident about its true reliability. In such cases, a more frequent, but slightly less informative form may yield grammars with higher confidence, and is selected as the base.

The results in (27) show that even this modest, indirect sensitivity to token frequency is enough to bias the learner towards selecting a more frequent form as the base when there is only a small difference in informativeness. This naturally raises the question: what is the trade-off between frequency and informativeness? How likely is the model to select the most frequent form over the most informative form as a base? Can it correctly predict the typological bias for analogical change to be based on

[20] In the case of Maori, the first part of this claim, at least, does seem to be true. Sanders (1990) found that over 70 percent of Maori verbs take either *-a* or *-tia* in the passive, and the choice of *-a* vs. *-tia* is itself somewhat predictable (Blevins 1994; de Lacy 2003). I have no information about the relative frequency of Maori verb forms to assess whether the second part of this claim also holds in the Maori case.

frequent forms? In the next section, I attempt to answer these questions by exploring the parameter space of the model.

7.4.2 The typological bias for frequent base forms

In the previous section, I showed that two factors play a role in favoring a less informative base form in Korean. First, the neutralization affects a small number of lexical items—only 1 or 2 percent of the nominal vocabulary. Second, the more informative forms have low token frequency (a greater than 50 percent difference). But how great must the frequency difference be, and how small the informativeness difference, in order for the model to choose a more frequent form over a more informative one?

In order to investigate this question, a series of artificial languages were constructed in which the most frequent member of the paradigm suffered from a neutralization (final obstruent devoicing), while the next most frequent member faithfully preserved all underlying contrasts. These languages varied along two dimensions: (1) the seriousness of the neutralization, and (2) the frequency difference between the most frequent paradigm member and the most informative one.

The seriousness of the neutralization was manipulated by varying the number of artificial stems that ended in obstruents: in a language with no final obstruents, final devoicing does not lead to any ambiguity, while in a language where half of the words end in voiced obstruents, final devoicing neutralizes contrasts in 50 percent of the vocabulary. The artificial languages were constructed using a stochastic algorithm, and the percentage of voiced and voiceless stem-final obstruents was then checked using Microsoft Excel, in order to ensure that they did indeed display the intended frequencies. Six degrees of seriousness of neutralization were considered, ranging from 0 percent to 50 percent at 10 percent intervals.

The second factor that was manipulated was the difference in token frequency between the most frequent form and the most informative form. This difference varied from nearly equal frequency (50 percent vs. 45 percent of all tokens) to extremely unequal frequencies (90 percent vs. 5 percent of all tokens). When crossed with the different degrees of neutralization, this yielded a total of 6×9, or forty-five artificial languages (one for each combination of neutralization vs. frequency difference).

In order to make the artificial languages resemble an actual language, pseudo-lexicons, containing 44,000 words each, were computer-generated. Since words in natural languages vary considerably in their token frequency, the lexical items in the artificial languages were assigned relative frequencies mirroring nouns in an actual language (Korean, as represented in the Sejong corpus). As a result, the artificial languages had a realistic profile of high- and low-frequency words. As an example, (28) shows the dozen most frequent words in the artificial language designed to have neutralizations in 50 percent of its lexicon.

(28) 50 percent Neutralization in the nominative (ambiguous forms in bold)

NOM	ACC	DAT	Frequency (per million wds.)
pip	pipa	pipi	9,289
lew	lewa	lewi	7,990
bagmoj	bagmoja	bagmoji	5,868
ran	rana	rani	5,346
zol	zola	zoli	4,624
vifdoj	vifdoja	vifdoji	4,295
lak	laga	lagi	3,753
lep	lepa	lepi	3,231
ras	rasa	rasi	3,009
zik	zika	ziki	2,928
etc.			

As with the Korean simulations above, input data for the learner was drawn from these artifical lexicons in Monte Carlo fashion. A learning trial consisted of 1,000 tokens of inflected forms, drawn randomly in proportion to the relative frequency of both the lexical item and also the inflection (see (26) above). On average, high-frequency lexical items were produced more often than low-frequency items, and frequent paradigm members were produced more often than infrequent ones. Continuing with the example from (28), when the frequency difference between nominatives and accusatives is 70 percent vs. 25 percent (the remaining 5 percent of the tokens being dative), a sample learning trial might look like (29). The token counts in (29) show that collectively, these twelve most frequent words were produced sixty-seven times; the remaining 933 tokens were forms of less frequent words.

(29) Most frequent words in a learning trial for the language in (28) (NOM:ACC:DAT ratio = 70:25:5)

NOM	NOM tokens	ACC	ACC tokens	DAT	DAT tokens
pip	14	pipa	1	pipi	1
lew	6	lewa	6	lewi	1
bagmoj	6	bagmoja	2	bagmoji	1
ran	5	rana	1	rani	0
zol	3	zola	1	zoli	1
vifdoj	5	vifdoja	0	vifdoji	0
lak	1	laga	0	lagi	0
lep	2	lepa	0	lepi	0
ras	0	rasa	1	rasi	0
zik	1	zika	3	ziki	0
mos	3	moza	0	mozi	0
mut	1	muda	1	mudi	0
etc.					

Ten learning trials were carried out for each artificial language, in order to ensure that the results were not distorted by a particularly deviant random sample. As for

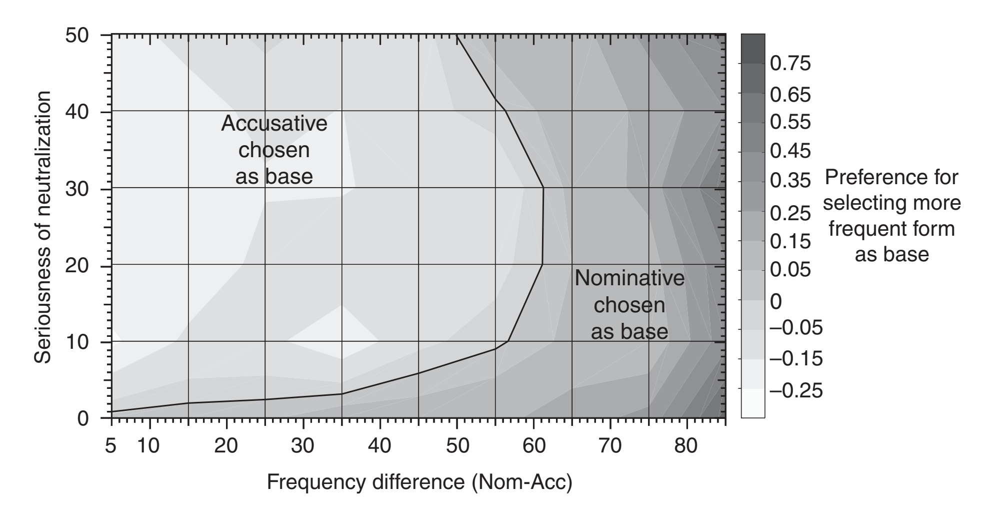


FIGURE 7.4. Trade-off between informativeness and frequency

Korean, accuracy and confidence were calculated for each paradigm member in each trial, and the results for each language were averaged across each of the ten trials.

The outcome of these simulations is summarized in Figure 7.4. In the graph, the two axes represent the two competing factors that can affect reliability of grammars: the seriousness of the neutralization, and the frequency difference between the paradigm members. The darkness of the shading indicates the degree to which the model prefers the more frequent form (the nominative) as the base; the spot at which this preference becomes positive (nominative is actually chosen) is indicated with a solid line. We see that although in all of these simulations the less frequent accusative form was more revealing about phonological contrasts, the model nonetheless has more confidence in the nominative form a fair proportion of the time. In particular, when the neutralization affects just a few words (seriousness is low), or when the frequency difference is great, the model's confidence in the accusative form suffers, and the nominative is chosen as the base.[21] What this graph shows is that even when the more frequent paradigm member is less informative, it is chosen as the base under certain conditions: in particular, when the frequency difference is great, or when the seriousness of the neutralization involved is below a certain threshhold (roughly, affecting less than 5 to 10 percent of the lexicon). In the previous section, we saw that Korean fits both of these criteria: the coda neutralizations that occur in

[21] The chart is bumpy rather than a perfectly smooth contour because it is based on randomly sampled input data, with statistical properties that varied slightly from the controlled properties of the artificial languages. Such anomalies are particularly noticeable towards the right side of the graph, where the frequency difference is great, and the probability of producing accusative forms is low. Such sampling fluctuations also appear to be responsible for the curious rounding at the top of the chart at a frequency difference of 50–60 percent (i.e., the nominative region juts to the left at the top of the chart), but more inquiry is needed in this area. It is expected that these irregularities would smooth out with more learning trials.

unsuffixed forms affect surprisingly few words, and there is a substantial frequency difference between unsuffixed forms and the remaining members of the paradigm, putting Korean at the extreme lower right of Figure 7.4.

The less frequent paradigm member, on the other hand, is chosen only when it is sufficiently frequent—even half as frequent may be sufficient—and when the benefit of choosing it is non-negligible. All of the typologically unusual cases discussed at the end of section 7.3 appear to have these properties. In the case of Yiddish, the frequency imbalance between the third- and first-person singular is most likely not as great as the difference between unsuffixed and nominative in Korean.

Although I know of no statistical counts for Yiddish, equivalent counts for Spanish (Bybee 1985: 71, and references cited therein) and German (Baayen, Piepenbrock, and van Rijn 1993) indicate that the ratio of first- to third-person singular tokens is typically about 1 to 1.8, compared to the 1 to 3.8 ratio found for Korean nouns. Moreover, the neutralizations affecting the third-person singular in Yiddish involve as much as 50 percent of the lexicon, placing it in the upper left region of the chart. Even less spoken frequency information is available to me about Lakhota or Latin, but these cases certainly involve more severe neutralizations than Korean, and seem to involve smaller frequency differences as well (see Albright 2002a: §6.2.1 for discussion).

In sum, the model predicts that even when the most frequent paradigm member is less revealing about lexical contrasts, it may nonetheless be chosen as the base a modest majority of the time. Crucially, we must bear in mind that this graph shows only half of the logically possible languages (those in which the more informative form is less frequent); when the most informative form also happens to be the most frequent one, it is *always* selected as the base. Taken together, this means that the general tendency for analogical change to favor more frequent paradigm members is correctly predicted, even though the model has no built-in bias to do so. As discussed at the end of the preceding section, the result follows simply from the fact that the learner is reluctant to trust generalizations based on too few data, and thus a slightly more informative base form can actually be less reliable if it is too infrequent.

7.5 CONCLUSION

In this chapter, I have argued that a synchronic model of paradigm acquisition can capture both language-particular and typological aspects of language change. On the one hand, the choice of base form, as well as the outcome of the eventual change (leveling vs. extension of alternations) can be seen to follow from structural properties of the language—in particular, the distribution of contrasts and the seriousness of neutralizations. On the other hand, the overall typological preference for analogy to extend less marked or more frequent forms is explained as a side effect of how learners

evaluate the seriousness of neutralizations, and the predictive power of potential base forms.

These predictions follow from a synchronically oriented model of language change, in which learners pay more attention to forms that are most helpful in predicting unknown forms, and analogical effects are rooted in this organization. The premise of this approach is that learners need to use limited information to learn how to produce and comprehend complete paradigms. In order to do this accurately and confidently, the learner must focus on those forms which permit the grammars with maximal confidence to be constructed for deriving the remainder of the paradigm. There are two sources of confidence under this model. First, rules are reliable when their inputs contain all of the necessary contrastive information to predict all of the surface forms; that is, when they do not suffer from neutralizations, forcing the grammar to probabilistically guess about outputs. Second, rules are reliable when they are general enough to have true predictive power. When rules are based on just a few examples, we cannot be confident that all future data will conform to the same generalizations that we have seen so far, whereas it is easy to trust rules that are based on ample data, even if it suffers from a few exceptions. Since less frequent members of the paradigm are sparsely attested in the learning data, learners are unable to confidently assess their predictive power, and are less likely to select them as bases.

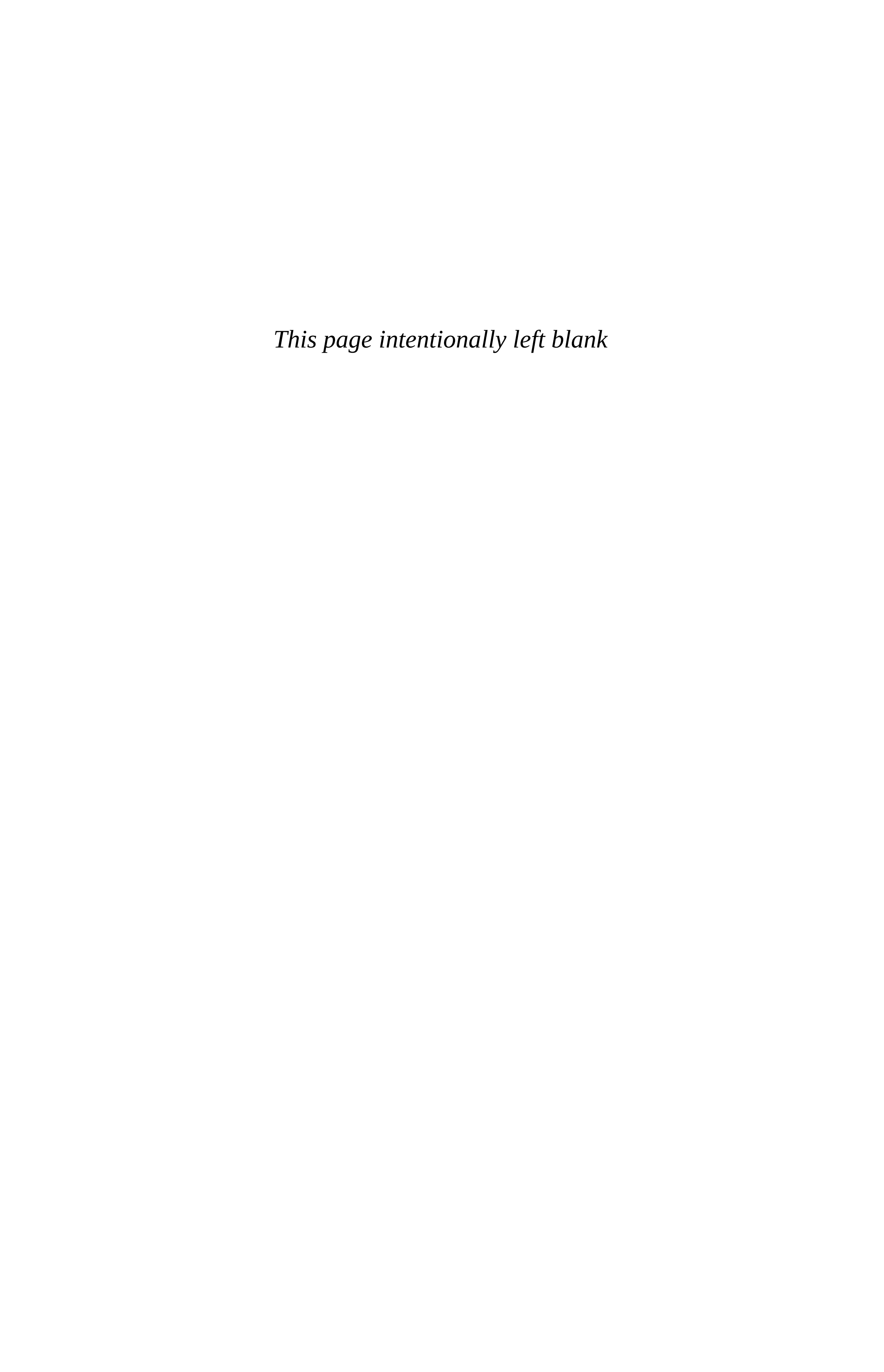
It should be emphasized that the base selection procedure that is proposed here is deterministic. Unlike a tendency-based approach, it is not based on statistical preferences or probabilities. This procedure is designed to generate a unique grammar for a given set of input data, and make predictions about likely changes.

An advantage to approaching analogical change in this way is that the model makes clearly testable predictions about both particular languages and the overall typology—so far, with promising results. Thus far, the focus has been on explaining typologically unusual cases of analogy, showing that they are sensible when viewed from the point of view of maintaining contrasts. An important result of this study, however, is the demonstration that predictability can be difficult to assess on the basis of schematic data, and that apparent neutralizations are often not as serious as they appear once one considers the number of words involved, correlations with other features, and so on. Furthermore, the Korean discussion highlights the fact that predictability can be influenced not only by the neutralizations involved, but also by the amount of data that is available to work with. The upshot is that the only true test of the model is one which takes all of these factors into account. If these results continue to hold with more detailed data and across a wider variety of cases, this approach has the potential to provide a more explanatory and predictive model of analogical change.

This page intentionally left blank

PART IV

Morphosyntactic Patterns: The Form of Grammatical Markers



This page intentionally left blank

8

Creating Economical Morphosyntactic Patterns in Language Change

Martin Haspelmath
Max Planck Institute for Evolutionary Anthropology

8.1 OVERVIEW

My starting point in this contribution is the observation that apparently the great majority of **universal morphosyntactic asymmetries** are economically motivated and thus exemplify the slogan "grammars code best what speakers do most" (Du Bois 1985: 363). By *morphosyntactic asymmetries* I refer to coding differences that do not express a meaning difference, e.g. the contrast between *the book* and *(*the) my book* (where the definite article *the* is impossible although from the meaning it would be expected to occur).

In sections 8.3–8.5 I will provide a substantial number of examples of economically motivated asymmetries. On the basis of these, I hypothesize that the pattern is even more general and that in fact the strong claim in (1) is correct.

(1) All universal morphosyntactic asymmetries can be explained on the basis of frequency asymmetries, i.e. they all show economic motivation: more frequent patterns are coded with less material.

In a second step, I want to examine the ways in which the economic motivation is implemented in languages through diachronic change. Economic motivation, like other types of functional motivation, needs to be interpreted in diachronic terms. From the point of view of the speakers' grammars, economical patterns are arbitrary (because speakers would be able to acquire non-economical patterns just as well), but the changes leading to them are not accidental; they are motivated by economy.

I am grateful for comments to the participants of the Berkeley workshop on Explaining Linguistic Universals, and to four anonymous reviewers. A version of this paper was also presented at a symposium on grammaticalization at the Freie Universität Berlin in November 2003 organized by Ekkehard König and Volker Gast.

So far, my study of the diachronic origins of economical patterns has not yielded very strong constraints, so the claim in (2) is perhaps not as surprising as the claim in (1).

(2) Diachronically, economical patterns arise by
  a. differential phonological reduction,
  b. differential inhibition of periphrasis/grammaticalization, or
  c. analogical change.
  'Differential morphosyntactic reduction', while logically perfectly possible, does not occur.

I will proceed by first explaining what I mean by "universal asymmetrical morphosyntactic patterns" (section 8.2). Then I discuss the notion of "economical coding" (section 8.3) and list eleven cases of "complementary expected associations" (section 8.4), and four cases of "non-complementary expected associations" (section 8.5). All these cases can be regarded as universal morphosyntactic asymmetries, and for all of them an economy-based explanation is plausible. In section 8.6 I discuss the diachronic origins of the fifteen economical patterns of sections 8.4–8.5.

8.2 UNIVERSAL ASYMMETRICAL MORPHOSYNTACTIC PATTERNS

By "asymmetrical patterns" I refer to situations in which one class of expressions behaves differently from another class of expressions without any apparent semantic reason.

(3) a. *a book of mine*
  b. *??the book of mine*
(4) a. *I saw you.*
  b. **You saw you.*
(5) a. *Who said what?*
  b. **What did who say?*
(6) a. *book/book-s*
  b. ??**stair/stairs*
(7) a. *sing/sang*
  b. *bring/*brang*
(8) a. *I'm interested in the picture.*
  b. **I'm surprised in the picture.*

In all these cases, the (b) expressions are perfectly interpretable, yet definite and indefinite NPs behave differently in possessive constructions (cf. (3)), disjoint and coreferential objects behave differently (cf. (4)), subjects and objects behave differently in multiple *wh*-questions (cf. (5)), *books* and *stairs* behave differently in the singular

(cf. (6)), and so on. So these behavior differences are initially surprising and require the linguist's attention. Other behavioral differences are unsurprising because they follow from the meaning of the expressions. For instance, *have to* is similar to *want to* in taking an infinitival complement (*I want to play/I have to play*), but behaves differently from it in not allowing a different-subject infinitival complement (*I want the kids to play* vs. **I have the kids to play*). But this is expected, because the latter sentence is simply uninterpretable on the obligation reading of *have to*. So I consider only those cases in which the asymmetry is surprising in the sense that symmetry would be possible and hence expected. In fact, in many cases there are languages which do not show the asymmetry in corresponding structures. For instance, Italian has no asymmetry in the equivalents of (3a–b) (*un mio libro/il mio libro*), and German has no asymmetry in the equivalents of (5a–b) (*Wer sagte was?/Was sagte wer?*).

Universal asymmetries are those that recur in language after language. Of the examples seen so far, (3) to (6) are of this type, but (7) and (8) are not, as far as we know: there is no cross-linguistic generalization under which the difference between *sing* and *bring*, and between *interested* and *surprised* could be subsumed. These seem to be language-particular idiosyncrasies, and no claims are made here about such non-universal, parochial patterns (which of course abound in all languages).

The universal patterns considered in this paper are all **implicational universals** in the classical Greenbergian sense (cf. Croft 2003: ch. 3). For instance, the difference between English and German with respect to multiple *wh*-questions can be described as an implication: if a language allows multiple *wh*-questions with fronting of the object and the subject remaining in situ, it also allows multiple *wh*-questions with fronting of the subject and the object remaining in situ. The reverse is not true, as English shows. Such universals are "typological generalizations" in Kiparsky's sense, not "true universals" (see Kiparsky, this volume), because (i) they arise as a consequence of tendencies of diachronic change (rather than constraining change), (ii) they are not assumed to derive from the innate cognitive code for grammar (often called "Universal Grammar"),[1] and (iii) they are not necessarily exceptionless. Notice, however, that one of Kiparsky's examples of a "true universal", the "D-hierarchy", is here regarded as a typological generalization (see section 8.4.4). In general, it seems that implicational universals are always "typological generalizations" in this sense, because proposals for Universal Grammar-derived implicational generalizations ("parametric effects") have been invariably unsuccessful (see Newmeyer 2004, 2005; Haspelmath 2008).

8.3 ECONOMICAL CODING

An expression shows **economical coding** compared to another expression if it is shorter (fewer words, fewer syllables, fewer segments) or otherwise requires less

[1] See Haspelmath (2004b) for the term "cognitive code", referring to the innate cognitive prerequisites for acquiring language.

articulatory effort (e.g. less suprasegmental prominence). Such economical coding is functionally motivated if it occurs with frequently expressed meanings, while related rarer meanings are coded with more articulatory effort. Four different kinds of economical coding asymmetry can be distinguished, and linguists often treat these in very different ways. My point here is that they all instantiate economical coding, and when they occur systematically, they require a unified explanation.

8.3.1 Frequent: zero/rare: overt[2]

In many cases, economical coding is manifested by a zero/non-zero contrast. Well-known examples are given in (9).

(9)		**Frequent expression**		**Rare expression**
a.	(i)	singular: *book-Ø*	(ii)	plural: *book-s*
b.	(i)	3rd person: Spanish *canta-Ø* 'sings'	(ii)	2nd person: *canta-s* 'you sing'
c.	(i)	present: *I Ø sing*	(ii)	future: *I will sing*

The overt non-zero element may be an affix (as in (9a–b)) or a free word (as in (9c)).

8.3.2 Frequent: shorter/rare: longer

Economical coding may also be manifested by a long/short contrast.

(10)		**Frequent expression**		**Rare expression**
a.	(i)	Tamil inanimate locative *-il*	(ii)	animate locative *-iṭam*
b.	(i)	Latin dative sg. *-ō/-ae/-ī*	(ii)	dative plural *-īs/-īs/-ibus*
c.	(i)	Russian 'middle' refl. *-sja*	(ii)	ordinary reflexive *sebja*

In (10c), the shorter expression is an affix, while the longer expression is a free form. This is another way in which short and long items often differ: long forms tend to show greater freedom and behavioral possibilities, while short forms tend to obey tighter restrictions. Since this difference cannot be related directly to economy, I will not discuss it in this paper.[3]

8.3.3 Frequent: straightforward/rare: roundabout

Often only the more frequently expressed meaning can be expressed in a straightforward way, while the rarer expression requires a roundabout construction:

[2] The cases on the right-hand side are not rare in an absolute sense (whatever that sense might be), and neither are the cases on the left-hand side frequent in an absolute sense. All that matters here is that there is a significant frequency difference between them.

[3] See Lehmann (1982 [1995]: ch. 4) for some discussion of the relation between shortness and lack of freedom.

(11)	Frequent expression	Rare expression
a.	(i) *Gabriel's friend*	(ii) *a friend **of** Gabriel's*
b.	(i) *I gave her it.* (British English)	(ii) *I gave it **to** Aisha.*
c.	(i) German *Ich will spielen.* 'I want to play.'	(ii) *Ich will, **dass** du spielst.* 'I want you to play.'
d.	(i) Modern Greek *Ton=íð a.* him=I.saw 'I saw him.'	(ii) *íð a **ton eaftó=mu**.* I.saw the self=my 'I saw myself.'

In (11aii), the possessor requires the extra preposition *of* and cannot occur in prenominal position. In (11bii), the full-NP recipient must occur with a preposition, so that the expression is longer and the two roles occur in different positions. In (11cii), the complement verb is finite and requires agreement with its subject, contrasting with the infinitive in the more frequent same-subject construction in (11ci). In (11dii), the reflexive pronoun does not occur in the preverbal pronominal slot, but in the postverbal slot of full-NP objects, i.e. despite its pronominal meaning it shows the behavior of a full NP.

While these cases are like those in (9) and (10) in that the rare expression is coded with more material, there are further differences in the constructions. But it is sometimes difficult to classify an asymmetric pattern, because an additional element may be regarded as itself constituting a constructional difference. Consider the pair in (12):

(12)	Frequent expression	Rare expression
	a. *sing!* (imperative 2nd pers.)	b. *let her sing!* (imperative 3rd pers.)

Here one would probably say that (12b) is a different construction and thus counts as "roundabout" like the (ii) cases in (11), but it is also possible to argue that this is a simple case of a zero/overt contrast of the type seen in (9). Thus, the boundaries between section 8.3.3 and sections 8.3.1 and 8.3.2 are not always clear, which is an additional motivation to treat these cases in the same way.

8.3.4 Frequent: existent/rare: nonexistent

In many cases, a language simply lacks a way of expressing the rarer meaning.

(13)	Frequent expression	Rare expression
	Tzutujil (Dayley 1985: 145)	
a.	(i) *w-ati7t* 'my grandmother'	(ii) **ati7t* 'grandmother'
b.	(i) *juyu7* 'mountain'	(ii) **w-juyu7* 'my mountain'
	Acehnese (Durie et al. 1994: 177–8)	
c.	(i) *Lôn-tém woe.* I-want return 'I want to return.'	(ii) **Lôn-tém droeneuh woe.* I-want you return 'I want you to return.'
d.	(i) *Who do you think that I met?*	(ii) **Who do you wonder why I met?*

At first glance, this case seems to be fundamentally different from the first three cases discussed in this section. However, it is clear that here, too, we observe "good" coding for frequently expressed meanings, in the sense that they can be coded straightforwardly, while the rarely expressed meanings are coded "badly" in an extreme sense: they cannot be expressed at all.

But when we look at this situation more closely, we realize that again, the dividing line between this case and the earlier clear cases of economical coding is not clear-cut. Languages are generally rich enough to render the meanings expressed by other languages, though sometimes significantly greater effort is required. So if a Tzutujil speaker is asked to translate 'grandmother' or 'my mountain' into Tzutujil (in an appropriate context), they would in all likelihood find a way of doing it. While Dayley's (1985) description is silent on this matter, one can easily imagine 'grandmother' being translated as 'a person's grandmother', and 'my mountain' as 'the mountain that belongs to me'. For Acehnese, Durie et al. (1994) suggest the sentence in (14) as a way of rendering 'I want you to return'.

(14) Acehnese (Durie et al. 1994: 177–8)
Lôn-lakèe droeneuh beu-neu-woe.
I-ask you HORT-you-return
"I ask you to return." (≈ I want you to return)

And if an English speaker wanted to say "Who do you wonder why I met?" in acceptable English, they could resort to a non-standard structure involving a resumptive pronoun (*Who do you wonder why I met them?*), or to a highly stilted expression such as *For whom is it the case that you wonder why I met him/her?*

Thus, under the assumption that selective ineffability does not exist (i.e. any language can be translated into any other language), the existent/nonexistent contrast reduces to the straightforward/roundabout contrast of section 8.3.3, though the degree of roundaboutness may be significantly higher in the examples in (13). While in the case of section 8.3.3, linguists perceive the (i) and (ii) cases as corresponding members of an opposition, this is not true in section 8.3.4, but of course we do not know whether such a difference exists for the speakers and how significant the difference is.

For this reason, I will consider cases like those in (13) as cases of economical coding, too. This goes beyond the common practice in the literature, but this is well-motivated because the cases of (13) all fall under the generalization in (1) (that is, if a meaning can be expressed directly in language A but only in a roundabout way in language B, then this meaning tends to be comparatively rare in language use).

Let us now look at a few of the asymmetries in somewhat greater detail, beginning with complementary expected associations.

8.4 ELEVEN COMPLEMENTARY EXPECTED ASSOCIATIONS

Morphosyntactic asymmetries often arise when two grammatical or semantic properties co-occur in one expression, and a particular value of property 1 is **typically associated** with a particular value of property 2. An example of this is the co-occurrence of *person* (= property 1) and *mood* (= property 2) in verb forms. In this case, the value "second" of person is typically associated with the value "imperative" of mood, i.e. the association "second/imperative" occurs more frequently (and is hence more expected) than others such as "third/imperative"; conversely, value "third" is the expected value of person with the value "indicative" of mood, so that the association "third/indicative" is more frequent than "second/indicative". This is thus an example of a **complementary expected association** (what Croft 1990: section 6.3 calls "complementary prototype").

In the following subsections, I will describe eleven such complementary expected associations, showing that they exhibit great similarities in their formal behavior across languages, and that this correlates in all cases with frequency asymmetries.

8.4.1 Person and mood

Our first example is the co-occurrence of person and mood. To simplify matters, we only look at second and third person, and at indicative and imperative mood, so that there are four possible associations of values, as shown in the four cells in Figure 8.1.

In this figure (and in the following figures in this section), the boxed cells are frequent (and hence expected) associations and show zero/short coding, while the non-boxed cells are rare/unexpected and show overt/longer coding. The overt coding elements in the non-boxed cells are highlighted by boldface. The cells contain minimally contrasting examples from a single language or two languages, but the figure (as well as all the figures that follow below) embodies two universal claims:

1. **Universal Frequency Asymmetry**: In all languages, what is identified as "expected association" here is more frequent than the corresponding "unexpected associations".
2. **Systematic Coding Asymmetry**: In all languages, the expected associations are at least as economically coded as the unexpected associations.

In order to make my claims watertight, I would have to provide evidence for both of these claims for each expected association. This is too ambitious a task, and to the extent that I do not provide the evidence, my contribution must be seen as providing a program for research rather than an empirical contribution. However, in most cases impressionistic observations give initial plausibility to my claims, and in some cases I

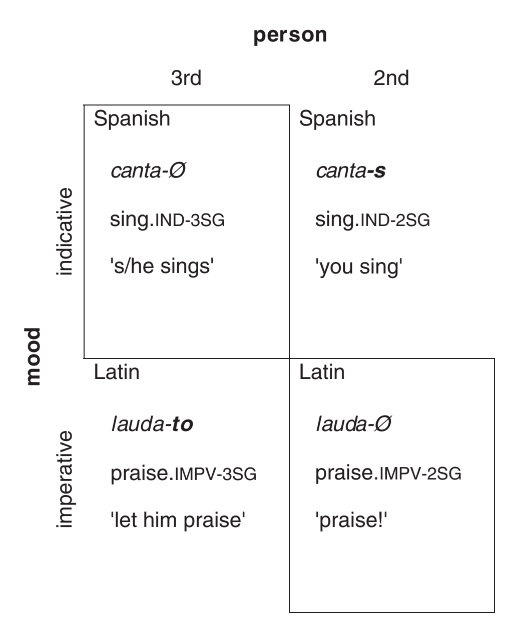


FIGURE 8.1.

will mention the results of very preliminary frequency counts that I did myself. These are mainly intended to show that the impressionistic observations do seem to have a basis in the facts and that doing systematic text counts of representative (i.e. colloquial) texts is a promising task.

The next two paragraphs describe the evidence that has been provided so far in the literature.

Evidence for frequency asymmetry: Greenberg (1966a: 47) provides text counts from three languages showing that the indicative is by far the most frequent mood, but he does not take person into account, and his texts do not represent colloquial speech and may therefore not be truly representative. However, it seems clear that in all languages, imperatives are most often second person (or first-person inclusive), because it is more efficient to address a command to the agent than to someone else.

Evidence for systematic coding asymmetry: Siewierska (2005) finds that out of 284 languages with verbal person marking, 103 have some zero-marked third-person forms. However, she does not compare these figures with zeroes in first- and second-person forms. Bybee (1985: 53) reports that 54 percent of her relevant languages (15 out of 28) use a zero for third-person subject agreement, while only 14 percent (4) use a zero for first-person, and only one language (Georgian) uses a zero for second-person indicative. For the imperative, the papers in Xrakovskij

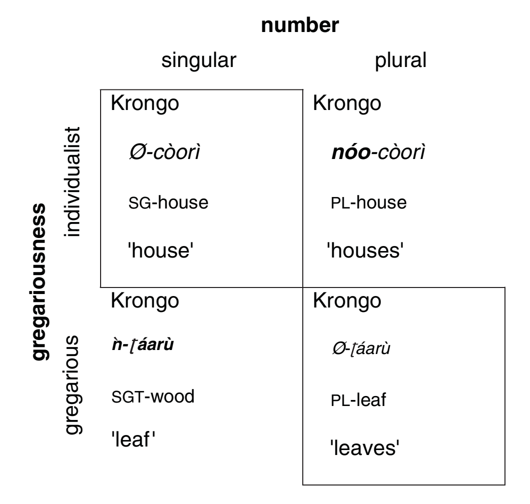


FIGURE 8.2.

(ed. 2001) show that there is a tendency for second-person imperatives to have zero-marking for person, and for third-person imperatives to have overt marking.

8.4.2 Number and gregariousness

The next example is the intersection of number and the lexical-semantic feature that I call here "gregariousness", i.e. this time we look at a grammatical-semantic property and a lexical-semantic property. "Individualist" nouns such as 'house' tend to be singular, and "gregarious" nouns such as 'leaf' tend to be plural. The effect on morphological coding can be seen in languages such as Krongo (Kadugli; Sudan; Reh 1985: 101ff.), which have both overt plural markers and singulative (= overtly coded singular) markers, as seen in Figure 8.2.

Evidence for frequency asymmetry: Greenberg (1966a: 32) shows with counts from four languages that the singular is overall more frequent than the plural, and this seems to be due to the fact that languages have more individualist nouns than gregarious nouns. Tiersma (1982) shows that some nouns, such as 'arm', 'horn', 'tooth', 'stocking', 'thorn', 'tear', 'leaf', tend to occur more frequently in the plural. It is these nouns that are called "gregarious" here.

Evidence for systematic coding asymmetry: Greenberg's (1963) universal number 35 states: "There is no language in which the plural does not have some nonzero allomorphs, whereas there are languages in which the singular is expressed only by zero." There are many languages in which all nouns behave like individualist nouns, but whenever there are singulative markers, they seem to occur with gregarious nouns. However, I know of no cross-linguistic studies of the lexical semantics of singulatives.

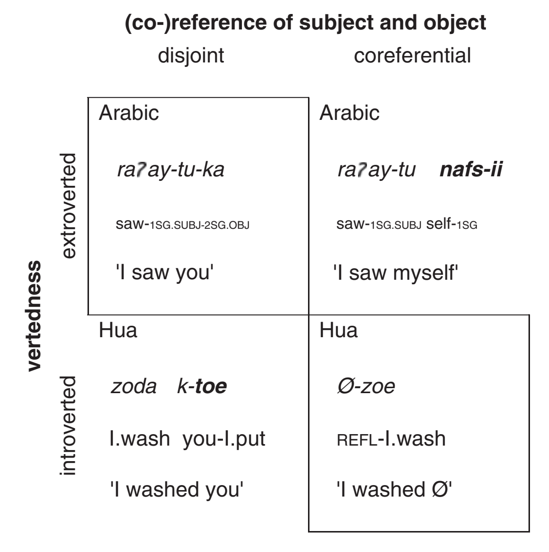


FIGURE 8.3.

8.4.3 (Co-)reference and "vertedness"

The frequency of coreference vs. disjoint reference of subject and object is affected by the lexical-semantic property of "vertedness". Some verbs such as 'see' are **extroverted**, i.e. the transitive action is typically directed toward another referent, whereas other verbs such as 'wash', 'shave', 'dress', 'defend' are **introverted**, i.e. the transitive action is typically directed toward the self (these terms are from Haiman 1983; see also Kemmer 1993; König and Siemund 1999). Examples, given in Figure 8.3, come from standard Arabic and Hua (Trans-New Guinea; PNG; Haiman 1983: 807).

Evidence for frequency asymmetry: already Faltz (1985: 7) characterized introverted verbs as denoting actions "that are commonly performed reflexively by people". In Haspelmath (2007a), I cite frequency data supporting this characterization.

Evidence for systematic coding asymmetry: Haiman (1983) and Kemmer (1993: 24–8) observe that introverted verbs are typically coded with shorter markers, and some languages have zero-coding for reflexive use with these verbs (e.g. English). Occasionally the disjoint (non-reflexive) use of these verbs requires a special overt marker, as in Hua (cf. Figure 8.3). The ordinary reflexive pronoun used with extroverted verbs is usually at least as long as the non-reflexive pronouns (cf. Faltz 1985; Haspelmath 2007a), and the situation of Arabic (cf. Figure 8.3), where disjoint pronouns are affixes but coreferential pronouns are full NPs, is quite typical.

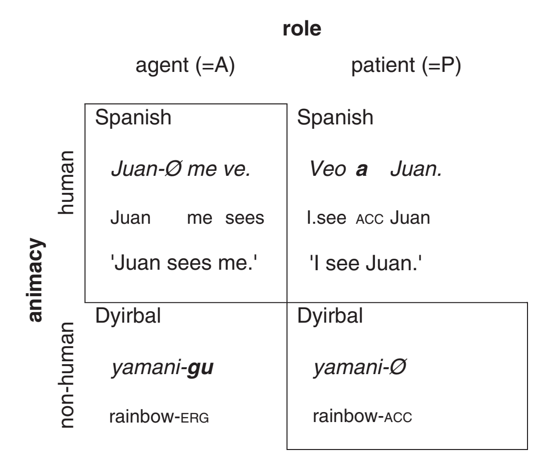


FIGURE 8.4.

8.4.4 Role and animacy

It is widely known that there is an interesting relationship between role and animacy: agents tend to be animate, and patients tend to be inanimate. The examples in Figure 8.4 are from Spanish and from Dyirbal (Pama-Nyungan; Queensland, Australia; Dixon 1972: 42).

Evidence for frequency asymmetry: Comrie (1989: 128) claims that "...the most natural kind of transitive construction is one where the A is high in animacy and definiteness, and the P is lower in animacy and definiteness", but what does "natural" mean here? It must mean the same as "expected", and the expectation derives from the frequency asymmetry. Jäger (2004) provides frequency data that fully support this.

Evidence for systematic coding asymmetry: the tendency for animate objects to be marked with more material ("differential object marking") has often been observed (Blansitt 1973; Silverstein 1976; Bossong 1985, 1998; Comrie 1989: ch. 6; Lazard 2001; Aissen 2003) and need not be discussed further. The tendency for inanimate subjects to be marked in a special way has been less widely discussed and is less widely observed, but to the extent that "differential subject marking" is observed, it also shows a systematic coding asymmetry (see e.g. Dixon 1994: 83–97).

8.4.5 Possessedness and alienability

In many languages, "inalienable" nouns (mainly body-part and kinship terms) and "alienable" nouns (all others) behave differently in possessive constructions. It used to

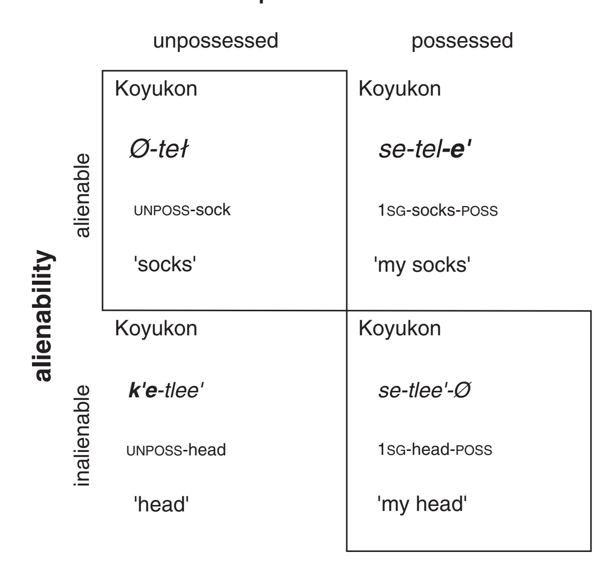


FIGURE 8.5.

be thought that this difference has to do with the different conceptualization of types of possession in these languages, but Nichols (1988: 579) pointed out that the different behavior of inalienables simply has to do with the fact that they occur more often as possessed nouns than the alienables. The examples in Figure 8.5 are from Koyukon (Athabaskan; Alaska; Thompson 1996).

Evidence for frequency asymmetry: I know of no frequency counts in the literature, but again it is quite easy to do such counts on texts. The figures in Table 8.1 are based on the German Goethe-Corpus of the COSMAS database[4] and show a few person-denoting nouns, occupational terms, and kinship terms. As expected, occupational terms are hardly ever possessed, while kinship terms are very often possessed. (There are relatively many cases of overtly unpossessed kinship terms because possessive pronouns may be left implicit in German, unlike in English.)

Evidence for systematic coding asymmetry: Nichols (1988: 579) observes that "the possessive affixes used on the closed ('inalienable') set of nouns are typically shorter, [and] involve fewer morphemes than the open class . . . ". Some languages simply lack unpossessed forms of inalienable nouns, and in some languages some alienable nouns are unpossessible (see the discussion of Tzutujil in section 8.3), but all languages allow inalienables to be possessed, and all languages allow alienables to be unpossessed.

[4] Institut für deutsche Sprache, Mannheim (http://corpora.ids-mannheim.de/~cosmas/).

TABLE 8.1.

		Unpossessed	Possessed
Alienable	*Gärtner* 'gardener'	24	0
	Jäger 'hunter'	48	2
	Pfarrer 'priest'	12	0
Inalienable	*Schwester* 'sister'	32	58
	Tante 'aunt'	47	22
	Tochter 'daughter'	46	53

8.4.6 Pragmatic function and semantic class

Most people would readily agree that nouns are predisposed for referring and verbs for predicating, but of course it is also possible to refer with verb-based expressions (e.g. nominalizations), and to predicate nouns (by using a copula), as in Figure 8.6. The predisposition again must be interpreted in terms of frequency (cf. Croft 1991: ch. 2): thing-denoting words more often have the pragmatic function of reference, and action-denoting words more often have the pragmatic function of predication.

Evidence for frequency asymmetry: Croft (1991: section 2.5) presents the results from text counts in four languages (Quiché, Nguna, Soddo, Ute), showing that thing-words are most often used for referring, and action-words are most often used for predicating.

Evidence for systematic coding asymmetry: Croft (1991: section 2.3) shows that across languages, overt "function-indicating morphosyntax" (such as copulas and nominalizers) tends to occur on the less preferred associations of semantic class (action/thing) and pragmatic function (reference/predication). Thing-words hardly ever require an overt nominalizer, and action-words rarely require an overt copula.

pragmatic function
reference
predication
semantic class
thing
English
book-Ø
(no nominalizer)
English
***is** a book*
action
English
*buy-**ing***
English
Ø buys
(no copula)

FIGURE 8.6.

FIGURE 8.7.

8.4.7 (Co-)reference and complement-taking verb class

Different complement-taking verbs have different preferences with regard to coreference of the main-clause subject with the subordinate subject. Verbs like 'want' have a strong preference for same-subject complements, while verbs like 'say' may have the opposite preference, for different-subject complements. The examples in Figure 8.7 are from Fongbe (Kwa, Niger-Congo; Togo; Lefebvre and Brousseau 2002: 78–82) and German.

Evidence for frequency asymmetry: I know of no evidence for complements of 'say' or other utterance predicates, but I did text counts for 'want' complements for two languages, Italian (Table 8.2) and Modern Greek (Table 8.3) (from Haspelmath 1999c). So clearly, same-subject complements are far more frequent than different-subject complements with 'want'. The Modern Greek evidence is particularly interesting because unlike German and Italian, Greek does not have two different constructions, i.e., it does not show any coding asymmetry. Thus, the frequency asymmetry cannot be due to the coding asymmetry.

Evidence for systematic coding asymmetry: in Haspelmath (1999c), I studied same-subject and different-subject complements of 'want' in fifty languages, and

TABLE 8.2.

Text frequency:	forms of *volere* 'want'	509	100%
(**Italian**)	same-subject	444	87%
	different-subject	65	13%

Source: Alessandro Manzoni, *I promessi sposi*, 1840–42 (Letteratura Italiana Zanichelli (LIZ) on CD-ROM).

TABLE 8.3.

Text frequency:	forms of *thélo* 'want'	43	100%
(**M. Greek**)	same-subject	38	88%
	different-subject	5	12%

Source: Kóstas Tzamális, *Stin Athína tu Periklí*, Athen: Estía/Kollaru, 22–122.

concluded that different-subject complement constructions are at least as formally complex as same-subject complements in all languages. Some languages simply lack different-subject complements of 'want' (cf. the discussion of Acehnese in section 8.3 above).

8.4.8 Possessedness and definiteness

In Haspelmath (1999a), I pointed out that possessed noun phrases have a greater chance to be definite than unpossessed noun phrases, so that definite articles are relatively redundant in possessed NPs, while indefinite articles are relatively redundant in unpossessed NPs, as in Figure 8.8.

Evidence for frequency asymmetry: see Haspelmath (1999a) for text counts in three languages.

Evidence for systematic coding asymmetry: in Haspelmath (1999a), I provided cross-linguistic evidence showing that there is a tendency for the definite article to be omitted in possessed NPs. The pattern found in English is attested (in various guises) in languages of different families.

possessedness

definiteness	possessed	unpossessed
definite	English *my Ø book* (no definite article)	English ***the*** *book*
indefinite	Hebrew ***eħad mi-****sfar-ay* one of-books-1SG 'one of my books'	Hebrew *sefer Ø* (no indefinite article)

FIGURE 8.8.

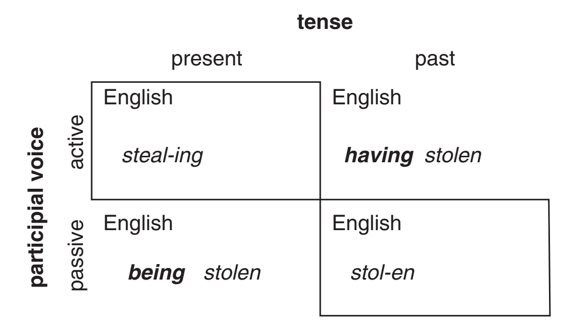


FIGURE 8.9.

8.4.9 Tense and participial voice

Comrie (1981) pointed out that there is a semantic affinity between perfect aspect and passive orientation, and we can expect this to be reflected in the frequency with which active and passive perfects and non-perfects are used. In participles, the expectation is that present participles tend to have active orientation (cf. English *stealing*), while past/perfect participles tend to have passive orientation (cf. English *stolen*), as illustrated in Figure 8.9.

Evidence for frequency asymmetry: in Haspelmath (1994), I did a text count of a language that shows no asymmetry in its system of participles (Lezgian; Nakh-Daghestanian; cf. Haspelmath 1993a) and showed that perfect and past participles more often show object-orientation than imperfective and habitual participles.

Evidence for systematic coding asymmetry: in Haspelmath (1994), I looked at systems of participles in a range of languages and found that when there are asymmetries in the systems, passive participles tend to be associated with perfect or past-tense interpretation, and active participles tend to be associated with present-tense or imperfective interpretation. Past active participles and present passive participles tend to be nonexistent, or (as in English) have to be expressed in more complex ways.

8.4.10 Sex and typical occupations

It is well known that occupational terms tend to show coding asymmetries with respect to sex, in such a way that male terms have no particular marking, while female terms exhibit a special affix. See Figure 8.10.

Evidence for frequency asymmetry: I know of no text counts, but a cursory glance at a frequency dictionary of English confirms the expectation that male occupational terms are generally more frequent than female terms (e.g. *king* 176, *queen* 80; *actor* 36,

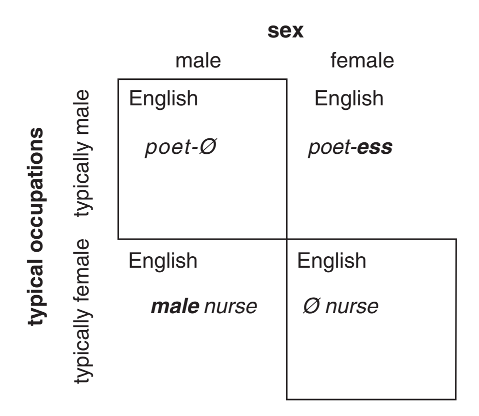


FIGURE 8.10.

actress 13; *policeman* 34; *policewoman* <10; Leech et al. 2001). Of course, this does not apply to typically female occupations, but at least in pre-twentieth-century societies, when most of the relevant linguistic patterns were formed, most occupations were far more often held by men than by women. And some typically female occupations do show the opposite coding pattern.

Evidence for systematic coding asymmetry: I know of no systematic cross-linguistic study of male and female occupational terms, but there seems to be little doubt that the English (and more generally, European) pattern is systematic across languages.

8.4.11 Transitivity and type of causation

Among situations that can alternatively be construed as occurring spontaneously and as occurring through external causation, two broad types can be distinguished (Haspelmath 1993b): **costly events**, for which it is more likely that the event is seen as caused by an agent (e.g. 'split', 'break', 'open', 'close', 'gather'), and **automatic events**, for which it is more likely that the event is seen as occurring spontaneously (e.g. 'boil', 'freeze', 'dry', 'sink', 'melt'). See Figure 8.11.

Evidence for frequency asymmetry: Wright (2001: 127–128) lists frequency figures for twenty-eight change-of-state verbs in English and finds significant differences between different English verbs. Verbs denoting events that I would describe as "automatic" (such as *corrode*, *ferment*, *dry*, *freeze*, *melt*) show a relatively low proportion of transitive occurrences, while verbs that I would describe as "costly" (such as *break*, *open*, *bake*) show a relatively high proportion of transitive occurrences.

type of causation	**transitivity**	transitive	intransitive
	costly events	Japanese *sak-Ø-u* 'split (tr.)'	Japanese *sak-**e**-ru* 'split (intr.)'
	automatic events	Japanese *kawak-**as**-u* 'dry (tr.)'	Japanese *kawak-Ø-u* 'dry (intr.)'

FIGURE 8.11.

Evidence for systematic coding asymmetry: evidence from twenty-one languages was presented in Haspelmath (1993b), showing that there is a tendency for costly events to be coded by transitive/anticausative alternations (i.e. alternations with overt coding of the intransitive verb), and for automatic events to be coded by causative/intransitive alternations (i.e. alternations with overt coding of the transitive verb). There are also tendencies within particular languages to be either uniformly causativizing or uniformly anticausativizing, but the correlation between the coding types and the types of verbs can be seen despite these language-particular differences.

8.5 NON-COMPLEMENTARY EXPECTED PATTERNS

Not all frequency asymmetries occur in complementary expected associations. Four such cases are presented in this section.

8.5.1 Agent–Patient direct/inverse

Quite a few languages require special coding effort for situations in which the person scale (1/2 > 3) is not aligned optimally with the role scale (Agent > Patient). These patterns are often called "inverse patterns". Relevant examples are given in Figure 8.12.

Evidence for frequency asymmetry: I know of no text counts, so there is no concrete evidence yet for a frequency asymmetry.

Evidence for systematic coding asymmetry: see Zúñiga (2006) for a recent overview of inverse systems in languages of the Americas.

direct (1 –> 3, etc.) inverse (3 –> 1, etc.)

Nocte	Nocte
nga-ma ate hetho-Ø-ang	*ate-ma nga-nang hetho-**h**-ang*
I-ERG he teach-DIRECT-1SG	he-ERG I-ACC teach-INV-1SG
'I will teach him.'	'He will teach me.'

(Nocte: Tibeto-Burman; India; DeLancey 1981: 641)

FIGURE 8.12.

8.5.2 Recipient–Theme direct/inverse

Although this is less well known, inverse-like patterns are also found in ditransitive constructions. In languages with bound object pronouns for both Recipient and Theme, it is often the case that bound object pronoun combinations are not possible when the situation is "inverse", i.e., there is no alignment of the person scale (1/2 > 3) with the role scale (Recipient > Theme). A better-known example of this is French, as in Figure 8.13.

Evidence for frequency asymmetry: Haspelmath (2004a) presents some evidence from a German corpus that the inverse combinations are significantly rarer than the direct combinations.

Evidence for systematic coding asymmetry: Haspelmath (2004a) formulates the Ditransitive Person-Role Constraint and provides evidence for it from a wide variety of languages. No counter-examples have been found.

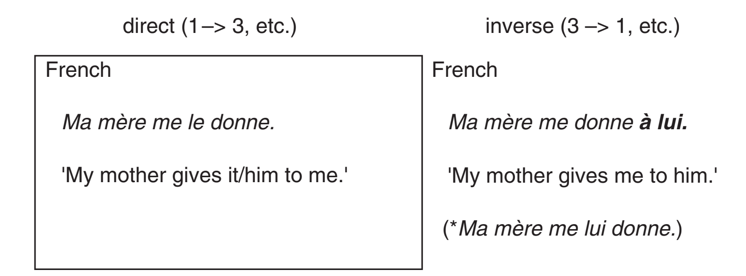


FIGURE 8.13.

8.5.3 Theme and instrument relativization

The famous Keenan-Comrie accessibility hierarchy for relativization is also an instance of economical coding. The grammatical relations that are "inaccessible to

semantic role

theme	instrument
Dyirbal *ŋuma [banaga-ŋu] yabu-ŋgu* father.ABS return-REL mother-ERG *buran.* see-NONFUT 'Mother saw father who was returning.'	Dyirbal *yugu [ŋuma-ŋgu balgal-**ma**-ŋu yabu-gu]* stick.ABS father.ERG hit-APPL-REL mother-DAT *jaja-ŋgu bura-n.* child-ERG see-NONFUT 'The child saw the stick that father used to hit mother.'

FIGURE 8.14.

relativization" in a given language are not inaccessible in an absolute sense, because there is usually a way to get around the constraints by employing a more complex construction. Most famously, a grammatical-relation-changing affix on the verb can promote a less accessible relation to a more accessible one. For example, in (15) relativization on the instrument is not possible, but apart from the ordinary construction as in (16a), Dyirbal also allows the applicative as in (16b), in which the instrument occurs as a core argument in the absolutive case (Dixon 1994: 169–171).

(15) **yugu [yabu ŋuma-ŋgu balga-ŋu] jaja-ŋgu bura-n.*
stick.ABS mother.ABS father.ERG hit-REL child-ERG see-NONFUT
'The child saw the stick with which father hit mother.'

(16) a. *yabu ŋuma-ŋgu balga-n yugu-ŋgu.*
mother.ABS father-ERG hit-NONFUT stick-INSTR
'Father hit mother with a stick.'
b. *yugu ŋuma-ŋgu balgal-ma-n yabu-gu.*
stick.ABS father-ERG hit-APPL-NONFUT mother-DAT
'Father hit mother with a stick.'

The instrument can be relativized after applicativization, as shown in Figure 8.14.

Evidence for frequency asymmetry: the Keenan-Comrie hierarchy has not in general been presented as a hierarchy of increasing rarity of relativization, but I strongly suspect that this is what underlies it.

Evidence for systematic coding asymmetry: Keenan and Comrie (1977) have provided ample evidence for the coding asymmetry. They do not in general say which strategies languages employ when the speakers want to relativize on a grammatical relation that cannot be relativized on directly, but clearly all languages employ strategies that are in some way more elaborate or complex.

subordinate-clause type	
'think that'	'wonder why'
English *Who do you think that I met?*	English *Who do you wonder why I met* ***them****?* *(*Who do you wonder why I met?)*

FIGURE 8.15.

8.5.4 'Think' vs. 'wonder' extraction

Linguists have not usually seen extraction restrictions in the context of frequency and economy, but I would like to suggest that the systematic grammatical asymmetries observed in this domain ultimately have to do with frequency, too.

Evidence for frequency asymmetry: since extraction phenomena occur very rarely in texts, it will be extremely difficult to do text counts, and indirect estimations of text frequency will have to suffice.

Evidence for systematic coding asymmetry: see Hawkins (1999, 2004: ch. 7) for some discussion of universals of extraction.

8.6 THE DIACHRONIC ORIGINS OF ECONOMICAL/WELL-CODED PATTERNS

Functionalists have sometimes been content with pointing out usage–grammar correspondences, because they confirm the expectation that "grammars code best what speakers do most" (Du Bois 1985: 363). But why are grammars well designed for the purpose of speaking and understanding? Why do they code best what speakers do most? Human beings are used to working with instruments that are well designed for their purposes, and in the case of human-made artifacts, such good design is not surprising because the creators' plan provides the link between the purpose and the structure of the instrument. For language, there is no such plan, so we need a theory that explains how language use and language structure are connected.

Following Bybee (1988b), Keller (1994), Kirby (1999), and related work (cf. also Haspelmath 1999b), I claim that diachronic change is the necessary link between patterns of language use and grammatical structures. Speakers do not intend to create well-designed grammars, but they behave purposefully and rationally in selecting from available variants and in creating new variants—they mostly opt for the most useful variants for their particular purposes. Through an invisible-hand process in language

change, the cumulative effect of many individuals' behavior leads to useful language structures (cf. Keller 1994).

So how do economical patterns arise in language change? There are two rather different routes by which this can happen: differential phonological reduction (section 8.6.1) and differential expansion of a new construction (section 8.6.2). Moreover, a minor route, morphological analogy, must also be recognized (section 8.6.3).

8.6.1 Differential phonological reduction

According to George Zipf (e.g. 1935, 1949), who was one of the first authors to emphasize the role of frequency for understanding linguistic structure, there is one main diachronic pathway by which frequency leads to shortness: **differential phonological reduction**. Frequent expressions tend to get shortened by phonological reduction more strongly than rare items. And of course there is no doubt that this is an important mechanism of change leading to economical patterns:[5] frequently occurring items are generally more predictable than rarely occurring items, so that hearers can decode the message even if it is not very carefully articulated. As a result, speakers tend to articulate them less carefully, so that they undergo faster phonological change than less frequent items. High frequency of use also leads to automatization, often implying affixation of formerly free function words (cf. Bybee 2003).[6]

Differential phonological reduction can be seen as responsible for a few of the economical patterns that we saw earlier. Some examples of phonological changes resulting in asymmetric coding are shown in (17)–(21).

(17) Person (section 8.4.1)[7]
Polish 3SG indicative forms became zero by special phonological reduction

[5] Historical linguists working in the Neogrammarian tradition have often ignored this insight, but there is massive evidence for it (see, e.g., the work of Witold Mańczak, such as Mańczak 1987).

[6] Bybee (2003) states that frequency-based automatization or routinization is not only responsible for increasing cohesion (= affixation, etc.), but also for phonological reduction. This is implausible, because frequently uttered and automatized expressions do not get reduced when they happen to be hard to predict. This is the case, for example, in the speech of individuals with regard to their names. I say my surname *Haspelmath* very often (at least much more often than people with other surnames), but I do not reduce it more than other people, because it has very low predictability. (I do reduce my signature, not because it is more predictable, but because there is no need for it to be recognizable at all.)

[7] Johanna Nichols (p.c.) doubts that the cases in (17) and (19) are due to differential phonological reduction, echoing some traditional views on these changes in the Neogrammarian-inspired literature. However, the alternative account in terms of frequency-induced reduction strikes me as much more plausible, especially for (17).

(cf. Proto-Slavic and the cognate language Russian):

	Proto-Slavic	Russian	Polish
1SG 'I write'	**pišǫ*	*pišu*	*piszę*
2SG 'you write'	**pišešĭ*	*pišeš'*	*piszesz*
3SG 's/he writes'	**pišetŭ*	*pišet*	*pisze-Ø*

(18) Number (section 8.4.2)
English singular of nouns became zero by special phonological reduction:
Old English *dæg/dagas* (> Modern English *day/days*)
< Proto-Germanic **dag-z*/**dag-ōs* (cf. Gothic *dags* 'day', *dagos* 'days')

(19) Reflexives (section 8.4.3)
Russian reduced reflexive pronoun: *-sja*, apparently derived by special phonological reduction from full reflexive pronoun *sebja*
(at the Proto-Slavic stage or even earlier)

(20) Alienability (section 8.4.5)
a. Old Italian Latin

moglia-ma	< *mulier mea*	'my wife' (inalienable)
fratel-to	< *fratellus tuus*	'your brother' (inalienable)
**terra-ma*	*(cf. terra mea)*	'my land' (alienable)

b. Nyulnyul (Nyulnyulan; northern Australia; McGregor 1996)

jan	*yil*	vs.	*nga-lirr*	(< *ngay lirr*)
I.OBL	dog		1SG-mouth	I mouth

'my dog' (alienable) 'my mouth' (inalienable)

(21) Complement clauses of 'want' (section 8.4.7)
English same-subject *wanna*, contrasting with different-subject *want to*
(*The reason I wanna come is Anna* vs. *The guest I want to come is Anna.*)

There may also be cases of differential phonological reduction of nominatives (section 8.4.4), but it is probably very difficult to find examples of phonological reduction leading to most of the other asymmetries. Zipf's diachronic mechanism of phonological reduction is thus less important in explaining grammatical asymmetries than one might have thought.

8.6.2 Differential expansion/inhibition of a new construction

Most cases of economical coding are due not to differential phonological reduction, but to differential expansion of a new, more complex construction (often called "periphrasis"). Such novel constructions typically make an existing meaning more transparent by including a special additional morpheme, and they are introduced when speakers want to call special attention to the relevant meaning, in particular when they want to express themselves in an especially clear

way (e.g. in situations of potential ambiguity). That is, initially these constructions are confined to highly unusual circumstances. A novel construction may then expand and become more frequent in an increasing number of new contexts, but it will be prevented from spreading to the contexts in which the relevant meanings occur most often. Such "inhibition of expansion" may occur for two reasons.

One reason is that the most frequently occurring combinations of meanings are the most deeply entrenched in the speakers' mental grammars and are thus unlikely to be replaced by innovations. Here we see the conserving effect of usage frequency.

But another reason, present in all cases of complementary expected associations of section 8.4, is the redundancy coming from the hearers' expectations: the expression of the meaning in question is redundant when it is typically associated with another meaning. Speakers know that hearers can predict the meaning that they want to express, so they are likely to economize and not use the novel, more explicit pattern.

Let us look at a few examples to see how this works concretely.

(i) Alienability splits (section 8.4.5)

In Classical Arabic, all nouns can take possessive affixes:

(22) *yad* 'hand' *kitaab* 'book'
yad-ii 'my hand' *kitaab-ii* 'my book', etc.

In Maltese, only inalienable nouns (body part terms/kinship terms) take possessive affixes; others occur in a periphrastic construction with *tiegħ-* 'of' (from *mtiegħ* < mataa ʕ'possession'):

(23) *id* 'hand' *ktieb* 'book'
id-i 'my hand' **ktieb-i* 'my book'
il-ktieb tiegħ-i (originally: 'the book my-possession')
the-book of-1SG

The novel construction involving the possessive noun did not expand to inalienable nouns: Maltese does not allow **l-id tiegħ-i* 'my hand'. There are two reasons for this: (i) the possessive suffixes are more entrenched in inalienable nouns and hence more resistant to loss; (ii) with inalienable nouns, the possessive semantics is more predictable and hence the new explicit construction is redundant.

Dahl and Koptjevskaja-Tamm (1998) make the strong claim that differential expansion is the only way in which an inalienability split can arise: "We suggest the generalization that an expanding possessive construction must encroach on the territory of pronominal possession for an alienability split to arise". However, we saw in (20a–b) above two examples in which such a split apparently

came about by differential phonological reduction, so this hypothesis cannot be maintained.

(ii) Differential Object Marking (section 8.4.4)

In Spanish, the preposition *a* was introduced at some point to mark patients (direct objects) in transitive clauses (this occurred by meaning change of the preposition, which formerly only expressed direction and various "dative" meanings). Initially the patient-marking *a* was limited to a few pronouns and nouns, but gradually it spread to all specific human NPs and became obligatory (*Veo a Juan* 'I see Juan'). However, it did not spread further than that: inanimate NPs do not take this preposition when occurring as patients (**Veo* ***a*** *la iglesia* 'I see the church'), i.e. the expansion of *a* was inhibited in the context in which patients/non-human NPs most typically occur. This was presumably (1) because verb–non-human NP combinations were more frequent and hence more resistant to innovation, and (2) because the patient role is more predictable with non-human NPs, so that speakers could afford not to use a specific marker in this environment.

(iii) Sex-specific affixes (section 8.4.10)

Imagine a language with no sex-indicating markers, perhaps Old Hungarian. In this language, an occupational term such as *orvos* 'doctor' can refer to males and females alike. At some stage a compounding strategy is introduced to specify the person's sex, e.g. *orvos-nő* 'female doctor' (*nő* 'woman'), perhaps also *orvos-férfi* 'male doctor' (*férfi* 'man'). Again, there are two potential reasons why only the *-nő*-compounds spread in the language (and in fact *-nő*-suffixation is now the main way of deriving sex-specific occupational terms). (1) Occupational terms with male meaning are highly entrenched and hence unlikely to be replaced by a new strategy. (2) For most occupations, the information that the person is male is redundant and only the less likely meaning combination needs to be indicated by speakers.

(iv) Recipient–theme inverse (section 8.5.2)

The French pattern in *me le (donne)* '(gives) it to me' is a direct continuation of the Latin *mihi illum*, i.e., it represents the old pattern. The roundabout construction involving the preposition *à* (*me donne à lui* 'gives me to him') is the innovated construction, which is now used with all full-NP recipients, but its expansion stopped before reaching the most frequent patterns: pronoun combinations showing the "direct" alignment of the person and role scales (Haspelmath 2004a). Again, two reasons can be given for this: (1) the pronoun combinations were deeply entrenched and hence more resistant

to change, and (2) the meanings of the frequent pronoun combinations were easier to predict than those of the rare combinations.

It seems that in different cases, the relevance of these two reasons for differential inhibition is different. Thus, the conserving effect of frequency seems to be very relevant for the inalienability split (subsection (i)), but less so for differential object marking (subsection (ii)), and hardly at all for sex-specific affixes (subsection (iii)). At present, I know of no way of assessing the respective roles of the two factors, and I leave this as an intriguing question for future research.

8.6.3 Excluded: differential morphosyntactic reduction

A priori, one could easily imagine another way in which economical patterns are created in language change: by **differential morphosyntactic reduction**. This would mean the omission of a particular morpheme or combination of morphemes which were originally present, but came to be omitted in a more frequent construction. However, I know of no evidence that such a change has ever occurred, and one may hypothesize that it is impossible.

Consider a concrete example whose history is not known (at least to me): in Ewe (Kwa, Niger-Congo), inalienable possession is indicated by simple juxtaposition, whereas alienable possession is indicated by a postposition *φé* (Ameka 1996: 791, 797).

(24)	Ewe	*kofí*	*srɔ̂*	vs.	*kofí*	*φé*	*awu*
		Kofi	spouse		Kofi	POSS	garment
		'Kofi's wife'			'Kofi's garment'		

If differential morphosyntactic reduction is impossible, we know that the present situation does not come from an earlier pattern where *φé* was present in both types of possession (*kofí φé srɔ̂/kofí φé awu*), and it was simply dropped from the construction where it was more redundant.

However, there are two rather serious problems with this proposed universal of diachronic change: (1) in practice, it is often difficult to exclude the possibility that the change was phonologically motivated and conditioned, (2) analogical change can create economical patterns, too (see next subsection), and we must make sure that this is taken into account in formulating the diachronic universal. Moreover, a reviewer mentions a possible counter-example: the change in many English varieties from *I've got* to *I got* (in the sense 'I have').[8] I do not know whether this could be explained by phonological reduction, so at the very least it illustrates the difficulty of assessing the validity of the claim that differential morphosyntactic reduction does not occur.

[8] Another potential counter-example mentioned by a reviewer is of a different nature: the frequent absence of the English complementizer *that* with the most frequent verbs that take *that*-complements (*say*, *know*, etc.). However, it is not clear that there was a stage at which complementizerless complement clauses with such verbs were impossible.

8.6.4 Analogical change can create economical patterns

An economical case-marking system may arise by **selectively preserving older markers**. For example, in the Old High German *n*-declension, animate and inanimate nouns alike had a distinction between nominative and accusative (cf. (25)).

		Old High German		Modern German	
(25)	NOM.SG	*affo*	*knoto*	*Affe*	*Knoten*
	ACC.SG	*affon*	*knoton*	*Affen*	*Knoten*
		'ape'	'knot'	'ape'	'knot'

Then the nominative–accusative distinction was lost in inanimate nouns, and in Modern German only animates preserve the zero-marking in the nominative. The resulting pattern shows differential object marking and thus economical coding, but it has come into existence via a different diachronic route than that discussed in subsection (ii) of 8.6.2 (cf. Haspelmath 2002: 245).

Another example comes from the history of French. Old French had the case- and number-marking pattern shown in (26).

(26) Old French

	SG	PL
NOM	*murs*	*mur*
ACC	*mur*	*murs*

In later French, the nominative–accusative distinction was given up, and those forms from both numbers were selectively preserved that lead to an economical pattern (SG *mur*, PL *murs*) (see Mayerthaler 1981: ch. 4 for some relevant discussion). Further related cases of analogical change are discussed in Bybee (1985: 54–56). (The approaches to analogy in Albright (this volume) and Garrett (this volume) do not seem to allow for the possibility that analogy may be motivated by the desire for an economical output; one wonders how they would deal with examples like those mentioned here.)

8.6.5 Is frequency really a causal factor?

A reviewer objected to the claim made here that frequency asymmetries can explain morphosyntactic asymmetries. Since this is an objection that I hear often and that still seems to be shared by many linguists, I quote the skeptical reviewer here:

> It seems to me more plausible that both the frequency and asymmetries are results, caused by something else. We say *male nurse* and (less often now) *lady doctor* because of cultural expectations—expectations that lead to *nurse* and *doctor* being used unmodified for females and males respectively. I can't see how frequency is in any sense the cause of the linguistic asymmetry.

The only problem with the idea that both frequency of use and the morphosyntactic asymmetries result from something else is that this something else cannot be identified. The only candidate that has been proposed, as far as I am aware, is "markedness", which is sometimes said to be responsible for frequency distributions (e.g. Mayerthaler 1981: 136–140). However, how "markedness" should influence frequency of use remains obscure, and it is relatively easy to show that, on the contrary, "markedness" effects of various sorts must be derived from frequency of use, so that the entire "markedness" concept can be abandoned (Haspelmath 2006).

Of course, the *male nurse/lady doctor* asymmetry is related to cultural expectations, but these imply linguistic expectations: since nurses tend to be women, words for female nurses occur more often than words for male nurses. As a result, saying *lady nurse* would be redundant and is avoided by speakers. One might be tempted to derive the redundancy directly from the cultural expectations, and in this particular case this would probably lead to the same results (though cultural expectations, or the real-world frequencies they presumably derive from, are much harder to measure than linguistic frequencies). But crucially, linguistic frequencies do not necessarily match real-world frequencies or non-linguistic expectations. Singulars are more frequent than plurals, though the world is hardly populated by more individuals than groups, and it would be hard to argue that there is a non-linguistic "expectation" that countable entities should occur singly. For whatever reason, people talk more about single entities than about pluralities, and this suffices to explain the facts of language. The important general point is that while linguistic frequency can usually be derived from other factors, these factors are quite heterogeneous. What matters to us grammarians is that the results of frequency are homogeneous, so that we can focus on linking frequency of use to economical patterns, leaving the extra-linguistic causes of frequency aside.

The skeptical reviewer also brings up the possibility of the direction of causation being the reverse: the availability of short coding could make speakers more likely to say a particular thing. In general this does not seem to be the case. Languages that have symmetrical coding patterns usually show the same kinds of frequency asymmetries as languages with asymmetrical coding. Thus:

- 'want' complements tend to be same-subject not only in asymmetric Italian, but also in symmetric Greek (see section 8.4.7);
- possessed NPs are preferentially definite not only in asymmetric English, but also in symmetric Italian (section 8.4.8; Haspelmath 1999a);
- active/present and passive/past participles are preferred not only in asymmetric English, but also in symmetric Lezgian (section 8.4.9; Haspelmath 1994);
- automatic events like 'freeze' tend to be more rarely used transitively not only in asymmetric Japanese, but also in symmetric English (section 8.4.11);
- inverse recipient/theme constructions ('show me to him') are rarer than direct constructions ('show him to me') not only in asymmetric French, but also in symmetric German (section 8.5.2; Haspelmath 2004a).

I conclude that frequency of use really is the relevant causal factor, and the reviewer's skepticism is unjustified.

8.7 CONCLUSION: THE RELATION BETWEEN DIACHRONY AND LANGUAGE UNIVERSALS

In this paper, I have made five main points:

1. A very large number of morphosyntactic implicational universals can be explained by invoking economic motivation (Haiman 1983):[9] more frequently used expressions are shorter than semantically similar, but more rarely used expressions, because they are more predictable.
2. Apparently all universal morphosyntactic asymmetries (in the sense of section 8.2) are economically motivated (see (1) above). This is a meta-universal, a universal about the explanation of universals.
3. Economical patterns are created by speakers in language use, and when innovative patterns spread through the community, they are manifested in the results of language change.
4. There are at least three different diachronic paths through which economical patterns arise: differential phonological reduction (section 8.6.1), differential expansion of a new construction (section 8.6.2), selective analogical change (section 8.6.4).
5. One obvious possible diachronic path does not seem to be well attested: differential morphosyntactic reduction (section 8.6.3). However, this generalization is problematic, because morphosyntactic reduction is not easy to differentiate from phonological reduction, and counter-examples have been noted.

If for the sake of the discussion we consider (27) as the most basic question of this volume,

(27) Do synchronic universals arise from universals of change, or do universals of change arise from synchronic constraints?

the answer given by the present paper is an intermediate one. While I do not deny that some synchronic universals derive from synchronic universals of the cognitive code (e.g., perhaps the generalization that grammatical rules do not include numerical specifications), the implicational universals considered here arise in language change. If we created an artificial language violating some of the universals, it would be learnable without problems, but it would be predicted that after a few centuries the language would give in to the pressure to change to more normal, economical patterns. In this regard, I am thus in agreement with Bybee (1988b, this

[9] Economic motivation is also the relevant factor for many of the phenomena that Haiman (1983) and others have attributed to "iconic motivation", as is shown in Haspelmath (2007b).

volume), Hopper (this volume), Blevins (this volume), and Garrett (this volume), who emphasize the importance of diachronic change for understanding synchronic universals.

However, I would argue that many diachronic changes cannot be understood without taking into account the result they lead to. It is not an accident that differential phonological reduction, differential expansion of a new construction, and selective analogical change may all lead to synchronic economical patterns (rather than, say, to counter-economical ones). Rather, the changes are motivated by economy, i.e., by the innovating speakers' desire to speak economically. This is not a synchronic grammatical constraint, or even a more general cognitive constraint. It is simply a constraint on any rational behavior.

Thus, with regard to the phenomena considered in this paper, the answer to (27) is: both. Implicational universals of the sort exemplified in sections 8.4–8.5 arise from universal tendencies of change toward economical patterns, and these universal diachronic tendencies themselves are motivated by the (synchronic) constraint that speech behavior should be rational and take both speakers' and hearers' needs into account.

9

On the Explanatory Value of Grammaticalization

Tania Kuteva and Bernd Heine
Heinrich-Heine-Universität Düsseldorf and Universität zu Köln

9.1 INTRODUCTION

Current research on linguistic change is faced with two contrasting hypotheses. On the one hand it is argued that language change is constrained by grammatical structure, or Universal Grammar, and can be explained by the way language is organized in the mind; let us call this the structure hypothesis (see Kiparsky, this volume). On the other hand it is maintained that since language structure is the product of processes that happened in the past, it can be explained with reference to these processes; we will refer to this as the grammaticalization hypothesis. It has been shown already that grammaticalization—as a theory[1]—has as its goal to describe the way grammatical forms arise and develop through space and time, and to explain why they are structured the way they are (Heine 2003). The main purpose of the present paper is to substantiate the grammaticalization hypothesis by looking at what is commonly viewed as "exceptional" phenomena in synchronic language structure. Moreover, we will argue that the grammaticalization hypothesis provides explanations that are essentially beyond the scope of the structure hypothesis.

There are different kinds of exception in grammar. Let us illustrate this in an area as basic in any grammar as personal deixis (person marking). First, an exception in a particular grammar can be a reflection of an exceptional sociohistorical circumstance. This has been proposed to be the case, for example, with the peculiar fact about Chinese Pidgin English that there is no form for the first-person plural *we*. Mühlhäusler and Harré (1990: 259) cite Forchheimer (1953: 12), who writes on

[1] Note that Heine (2003: 1) builds a case for a threefold distinction between (i) grammaticalization, which relates to specific linguistic phenomena, (ii) grammaticalization studies, which deal with the analysis of these phenomena, and (iii) grammaticalization theory, which proposes a descriptive and explanatory account of these phenomena.

Chinese Pidgin English: "Though I have found several languages where the word 'I' can also serve to express 'we', they all possess, besides that, a word for 'we'. The only exception is Chinese Pidgin English." Without going into details about the reliability of the empirical basis for such a statement, the explanation implied by Mühlhäusler and Harré (1990: 259) involves a reinforcement "by the peculiar social context in which Chinese Pidgin English developed". The absence of first-person plural *we* is taken to correlate with the lack of solidarity between the trade partners; a special significance is ascribed to the fact that Chinese Pidgin English was typically restricted to dyadic rather than group communication. In other words, since the social context is just a dyadic *you* (= second-person singular) versus *me* (= first-person singular) trading ("it's never *you* ***and*** *me together on the same side*, it's always *you* ***against*** *me*"), why bother about a group *we*—at least as a first-person plural inclusive category?

There is, however, another dimension of the discussion on exceptions in grammar: sometimes the same set of linguistic facts may constitute an exception for one theoretical model but a welcome fulfillment of an expectation for another; in other words, what appears to be an exception for one theoretical model turns out to be the rule for another model.

The data in (1) is an example from the personal deixis area again (Heine and Kuteva 2007). It involves a paradigm of forms which is ubiquitous in our everyday discourse in English; it exists practically under our noses.

(1) Standard English

	Singular number	Plural number
1st person:	I **was** sick	We *were* sick
2nd person:	You *were* sick (*You **was** sick)	You *were* sick
3rd person:	She **was** sick	They *were* sick

Here the paradigm of personal deixis exhibits a regular pattern of iconic number marking: the copula *was* is used with singular subject referents, and *were* with plural referents. However, there is one exception: the iconicity principle is violated in the second-person singular, which patterns with the plural forms rather than with the other singular forms. Note that the expected combination with *was* is not acceptable (*You **was** sick). Now, on theoretical models focusing on synchronic language structure, this exception is just a case of a so-called "irregularity", and irregularities are simply considered to be "facts of life", facts we have to live with and not bother about, and anyway facts like these are not considered "interesting": what matters are the regularities.

In one particular theory, however, the one we will be applying in this paper—that is, grammaticalization theory—facts like the latter are extremely interesting, and in need of explanation. The standpoint taken in this theory is that language structure is not necessarily regular, and that it is the frozen product of past processes of social and linguistic interaction. Thus the exceptional behavior of the number agreement pattern for second-person singular subject pronoun described above is readily explained as the

expected rule of the regular agreement pattern that obtained between the copula and the personal pronoun from which modern English second-person singular derived, that is, the form for the second-person plural. In other words, from a grammaticalization point of view, it is crucial to consider the following historical fact: the English pronoun *you*—which originally encoded second-person plural—has come to replace the earlier form for the second-person singular *thou* (nominative) and *thee* (accusative). Thus, like a number of other languages, English, too, has grammaticalized a particular rhetorical strategy—referred to as pluralization—based on a particular form of behavior in social interaction, whereby the speaker shows more respect to the addressee by pretending the latter is "more". While *you* has changed its erstwhile meaning, it has retained its number agreement pattern, and this trace of its earlier morphosyntactic characteristics is something which is expected within a grammaticalization theory framework.

It is this latter kind of exceptions that are the focus of our interest in the present paper. The theory which renders exceptions as rules—the "good guy" in our story—is grammaticalization theory. However, we will show that in order to fully reveal the explanatory potential of grammaticalization theory, it is necessary to move further from the version in which this theory has been used most of the time in the literature. For this purpose, we will employ the theoretical apparatus of grammaticalization in such a way that it encompasses both language-internal and contact-induced grammaticalization developments. Whereas language-internal grammaticalization has figured prominently in studies of grammatical language change in the past decades, contact-induced grammaticalization has not been given due recognition in the existing literature.

Taking into account both language-internal and contact-induced grammaticalization—we contend—enables the analyst of language to take advantage of the strong explanatory value of grammaticalization theory.

We will argue that from a grammaticalization theory perspective, what is the "exception" on a static, synchronic analysis readily becomes the "rule" of a developing grammar in both contact-related and non-contact-related situations.

9.2 THEORETICAL PRELIMINARIES: LANGUAGE-INTERNAL AND CONTACT-INDUCED GRAMMATICALIZATION

Grammaticalization is defined as the *development from lexical to grammatical forms, and from grammatical to even more grammatical forms*. Since the development of grammatical forms is not independent of the constructions to which they belong, the study of grammaticalization is also concerned with constructions, and with even larger



FIGURE 9.1. The traditional view: grammaticalization vs. language contact

discourse segments. In accordance with this definition, grammaticalization theory is concerned with the genesis and development of grammatical forms and constructions. Its primary goals are to describe how grammatical forms and constructions arise and develop through space and time and to explain why they are structured the way they are.

To date the dominant practice has been to study grammaticalization as an internally motivated process, that is, one that is motivated only by factors that are internal to a language and to the community of its speakers. With a few notable exceptions (see the volumes of the EUROTYP project, cf. König and Haspelmath 1999), the relevant literature abounds—indeed—with discussions on whether some specific grammatical change is due to grammaticalization or to contact-induced language change. That is, grammaticalization and language contact have been traditionally viewed as two mutually exclusive forces, or at least as forces that are independent of one another.

This traditional wisdom culminates in the view that grammaticalization, or language-internal change, is "natural", whereas contact-induced change, or language-external change, is "non-natural":

> [L]inguistic changes may come in two rather different types. Some forms of linguistic change may be relatively 'natural', in the sense that they are liable to occur in all linguistic systems, at all times, without external stimulus, because of the inherent nature of linguistic systems themselves—and it is here of course that the stability of the nature of human beings is relevant. Other types of linguistic change, on the other hand, may be relatively 'non-natural', in the sense that they take place mainly as the result of language contact. They are, that is, not due to the inherent nature of language systems, but to processes that take place in particular sociolinguistic situations. (Trudgill 1983: 102)

In our approach (Heine and Kuteva 2003, 2005; Kuteva forthcoming), language contact is not a factor working *against* grammaticalization; rather, it may work in conspiracy *with* grammaticalization. Hence, in addition to language-internal grammaticalization—which has been the object of investigation in numerous studies already—there is also contact-induced grammaticalization. Heine and Kuteva (2003) define contact-induced grammaticalization as a grammaticalization process that is due to the influence of one language on another (see figure 9.1).

On the basis of our study of language-contact phenomena and grammatical language change in the languages of the world (Heine and Kuteva 2002), we have shown (Heine and Kuteva 2005) that language-internal and contact-induced grammaticalization can be characterized in terms of the same set of criteria, with unidirectionality built into each of them.

Criteria of grammaticalization

a. extension (or context generalization): use in new contexts suggests new meanings;
b. desemanticization (or "semantic bleaching"), i.e., loss in meaning content;
c. decategorialization, i.e., loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms;
d. erosion (or "phonetic reduction"), i.e., loss in phonetic substance.

Moreover, there is abundant data to indicate that:

1. grammaticalization is a gradual step-by-step process which can be observed both in time (in the historical development of a language) and in space (in the synchronic, geographical-linguistic variation within a language);
2. as a grammaticalizing structure undergoes development from one stage to the next, there is often an overlap between the variants of that structure characteristic of two adjacent stages (Heine et al. 1991: 65–69).

9.3 TWO PUZZLES ABOUT DEFINITENESS MARKING IN EUROPE

9.3.1 The postposed definite article in Bulgarian

In some languages the definite article is preposed (e.g., German, English, French, Italian, Greek); in others it is postposed (Swedish, Icelandic, Armenian, Rumanian, Albanian, Bulgarian, Kurdish). The fact that the definite article can be either preposed or postposed is not unexpected from a grammaticalization point of view. Since definite articles develop from demonstrative attributes, they should occupy the position which the demonstrative occupies; in some languages the demonstrative is preposed, hence the definite article should also be preposed, in others the demonstrative is postposed, hence the definite article should also be postposed. Therefore, the correlation between the preposed demonstrative and the preposed definite article in English is not surprising.

(2) English
this/that table ⟵⟶ *the* table

By the same token, it will not be surprising if a language like To'aba'ita develops a postposed definite article since in this Oceanic language the demonstrative follows the noun, as in (3).

(3) To'aba'ita (Lichtenberk 1991: 493)
Si manga n- e toda- a thaari 'eri, ka ngata bii- a.
CL time REL- he.FACT meet- her girl that he.SEQ speak with- her
'When he met the girl, he spoke with her.'

Notice that in this example the English translation contains a definite article; experts on To'aba'ita, however, find that there is no grammaticalized definiteness marker in the language (Frantisek Lichtenberk, p.c.)

There exists at least one example which seems to contradict the grammaticalization scenario *demonstrative > definite article* because it involves a situation where the definite article does not occupy the same place as the demonstrative. This language is Bulgarian. The definite article in Bulgarian has the status of a suffix added to the nominal stem, as in (4).

(4) Bulgarian
masa- ta
table- the.F.SG
'the table'

In the literature on the historical development of Bulgarian it has been widely accepted that—just like in other languages with a definite article—the Bulgarian definite article can be traced back to the demonstrative. However, the Bulgarian demonstrative occupies a preposed position with respect to the nominal phrase, just like in English and German, so that 'this table' is rendered by the structure [demonstrative + noun], as in (5).

(5) Bulgarian
tazi masa
this.F.SG table.F.SG
'this table'

The immediate question that arises then is: how can we explain this violation of what we observed above, namely the grammatical typological correlation on the one hand, and the prediction based on grammaticalization theory, according to which preposed demonstrative attributes should give rise to preposed definite articles, on the other? In other words, why is the demonstrative "moved" from its preposed to its postposed place after the noun stem?

From the point of view of grammatical theories expecting demonstratives and determiners to appear in similar (or even the same) position in a syntactic structure, explaining this apparent contradiction is not easy. But, taking the perspective of grammaticalization theory, there is a straightforward account. At the time when the category of the definite article was taking shape (ninth to thirteenth centuries), the demonstrative could be used in a position after the noun (Mirčev 1963: 179), so that when the developing definite article had lost the deictic force of its historical source (the demonstrative), it found itself "frozen" in the same position in which that source could be used at the beginning of the grammaticalization process from demonstrative to definite article. The example in (6) illustrates the syntactic position of the demonstrative at earlier historical stages of Bulgarian; it comes from a text (*Narodno žitie na Ivan Rilski*) composed in the twelfth century and preserved in rewritings from the fifteenth century.

(6) Bulgarian (Mirčev 1978: 203; transliteration and glosses are ours)
I pridošă na MJASTO TO, ideže stoaše na kameni šty otc' Joann',
and came.they on PLACE THIS where stood on stone saint father John
molja boga. I pridošă naprasno să straxom' veliem' mnjašte
praying.to God and came.they all.of.a.sudden with intimidation great thinking
ustrašiti ego, xotjašte da bežit ot MJASTA TOGO štogo . . .
frighten him wanting to force.out from PLACE THIS.ACC sacred.ACC
'And they came to THIS PLACE, where the saint Father John stood on a stone, praying to God. And they came all of a sudden with many threats, thinking that they will frighten him, because they wanted to force him to leave THIS / THE sacred PLACE . . .'

It was only at a later historical time that the demonstrative acquired its present-day, fixed, preceding position with respect to the noun; at this later time the definite article was a well-established category already, with a fixed syntactic position within the nominal phrase.

Note that taking a grammaticalization approach to a seeming exception like the one under discussion here can not only "convert" this exception into an expected rule but it also enables us to reconstruct the genesis of the grammaticalized structure. Let us take a closer look at the early example of the demonstrative functioning as a definite article in (6) above. It involves a relative clause, i.e., a context where an "indexing" element (the demonstrative) is used in order to foreground information (the contents of the relative clause), and one nominal phrase with an adjective following the noun. In both cases, the demonstrative follows the noun and precedes the modifier, the latter being either a relative clause or an adjective, as schematized in (7).

(7) Pivot context for development of Bulgarian definite article
NOUN DEMONSTRATIVE RELATIVE CLAUSE / ADJECTIVE
place this where . . . / sacred
'the place where . . . ' / 'the sacred place'

In other words, it seems plausible to assume that the type of context where an erstwhile demonstrative element takes on the function of a definite article has to do with foregrounding information by modifying it with the help of either an adjective or a relative clause. That nominal modification may, indeed, be the trigger for a context-induced reinterpretation to bring about the rise of a new definiteness function is corroborated by the following diachronic fact about Old Church Slavonic: definiteness in Old Church Slavonic was marked—by the anaphoric demonstrative *ji* (M) / *ja* (F) / *je* (NEUT)—only on adjectives.

(8) Old Church Slavonic
vino novo- je
wine new- DEM
'the new wine'

This fact about Bulgarian, and the language it can be traced back to, Old Church Slavonic, is compatible with cross-linguistic data about the pattern of a noun phrase

containing adjectives as modifiers, where the definiteness marker is attached to the adjective only. Dahl observes:

> The existence of articles that mark adjectives only . . . indicates that the initial stages of the grammaticalization of definite articles may be restricted to NPs containing modifiers. As a possible explanation of such a development, consider the fact that an adjective or a relative clause (used restrictively) commonly singles out a subset within the set denoted by the head noun, contrasting it to its complement set. . . . Similar uses are plausible candidates for being the first step in the development of specific attributive articles. It is possible that they could later be extended into general definite articles, but as far as I know no such development has been properly documented. (Dahl 2004b: 152)

The account we have proposed of the postposing of the definite article in Bulgarian is compatible with an account whereby language contact on the Balkans must have also played at least an accelerating and/or a reinforcing role (see Breu 1994: 53). Indeed, a postposed definite article is what has traditionally been considered a characteristic of the "Balkanization" of the languages of the Balkan Sprachbund. That such a process of Balkanization must have taken place is supported by even just the presence of a definite article in Bulgarian—as well as in Macedonian, also a Balkan language—as contrasted to Slavic languages spoken outside the Balkans; the postposing of the definite article in these two languages makes the argument even stronger. A second argument in favor of language contact as a contributing factor comes from the fact that Rumanian is the only Romance language where the definite article is postposed as well as the only Romance language which is spoken in the Balkans. Thus Rumanian, like all other Romance languages, has, indeed, a full-fledged definite article. However, unlike the Western Romance languages and in accordance with what is found in other Balkan languages (except Greek), the article is postposed rather than preposed. Still, the Rumanian article is due to the same grammaticalization process as those of the Western Romance languages, being the result of a development from a distal demonstrative attribute (Latin *ille* M, *illa* F, *illud* N) to definite marker, e.g., Rumanian *omul* (< Latin *homine(m) illu*) 'the man' (Haarmann 1976: 85).

9.3.2 "Double determination" in Scandinavian

(i) Description of the "double determination" phenomenon

The second "puzzle" in the area of definiteness is the so called "double determination" in some Scandinavian language varieties. "Double determination" in Scandinavian involves the double marking of definiteness in Swedish (and most forms of Norwegian) as contrasted to other Scandinavian varieties (Danish and marginally in Norwegian). Generally speaking, there are two definite articles in Continental Scandinavian varieties (Danish, Norwegian, and Swedish): (i) a preposed article, homophonous with the demonstratives *den*, *det*, *de*, e.g., (9), and (ii) an article which is suffixed to the head noun, e.g., (10).

(9) Danish (Dahl 2004b: 147)
det store hus
DEM big house
'the big house'

(10) Danish (Dahl 2004b: 148)
hus- et
house- DEF
'the house'

Now, the preposed article is only used when the noun is preceded by an attribute. However, while in Danish and marginally in Norwegian the suffixed article is restrained whenever the prefixed article is used, in Swedish and most forms of Norwegian both articles are used in cases when the head noun is preceded by an attribute, as in (11).

(11) Swedish and most forms of Norwegian (Dahl 2004b: 147)
det stora hus- et
DEM big house- DEF
'the big house'

The "inconsistent" pattern exemplified in the above example is hard to reconcile with widely held beliefs (for a discussion of the widely held belief of a structural determiner position, for instance, see, Haspelmath 1999b: 228–231). In what follows we will show that from a grammaticalization point of view, there is nothing exceptional about it; rather, this kind of behavior is a natural result of both language-internal and contact-induced grammaticalization processes.

(ii) Explanation

Following the detailed description presented in Dahl (2004b), and employing an integrative model of grammaticalization whereby both language-internal and contact-induced phenomena are taken into account, we will show that the explanation of the exceptional situation referred to as "double determination" in Scandinavian is twofold. We will argue that the "double determination" phenomenon is the manifestation of an overlap stage of language-internal grammaticalization on the one hand, and a Buffer Zone in contact-induced grammaticalization, on the other. Such an explanation is also in accordance with an aspect of contact-induced grammaticalization we have articulated elsewhere (Heine and Kuteva 2006; Kuteva forthcoming), namely that contact-induced grammaticalization simultaneously implies both external and internal language change.

Overlap stage of language-internal grammaticalization

Swedish is typically cited as a paradigm example of a language with a postposed definite article.

(12) Swedish (Diessel 1999: 136)
Jag vill ha tillbaka bok- en som du lånade i förra veckan.
I want have back book- DEF REL you borrowed in last week
'I'd like the book back that you borrowed last week'

Moreover, the development of the postposed definite article is well attested in the history of the language: it derived from a demonstrative, which, like the Bulgarian demonstrative, must have been used in postposition—we argue—to the noun at an earlier stage of Scandinavian. Hence its present-day suffixal status.

As often observed by students of grammaticalization, however, the existing affixal expression of a grammatical category may start fading out (and ultimately get reduced to zero) and it may well be replaced by a new form, encoding the same category. Such a renewal of grammatical structures usually involves a free (non-bound) form, which over time may or may not acquire an affixal status again, and the process may repeat itself more than once in the lifetime of a language (on the so-called cyclic nature of grammaticalization, see Heine et al. 1991: ch. 8). One of the characteristics of this renewal of a grammaticalization process, however, is that there usually exists an overlap stage. This overlap may be twofold (see also Kuteva forthcoming). The first kind of overlap—also the one most often discussed in the literature—involves the existence of the same form, that is, the grammaticalizing element, which can potentially encode either the historically earlier meaning A or the historically later meaning B. This kind of overlap can be schematically represented as $A \to A/B \to B$. The overlap stage is hence one of ambiguity between the newly emerging meaning and the earlier original meaning.

The second kind of overlap, however, involves the coexistence of the older form a with the newly emerging form b within an existing grammatical structure which remains essentially the same from the point of view of function, which can be schematized as $a \to a + b \to b$. Thus the Swedish example of double determination above is a manifestation of a typical overlap stage of grammaticalization where the historically earlier, suffixal definite article and the newly arisen, preposed definite article *den* coexist. That the latter is a form still undergoing a grammaticalization process is indicated by the fact that without the adjective in the nominal phrase, the free, preposed definite article tends to be interpreted as a demonstrative.

In other words, "double determination" in Swedish is a manifestation of a natural—from a grammaticalization point of view—overlap stage where the postposed article reflects an earlier historical stage, and the preposed article a more recent one.

Buffer Zone in contact-induced grammaticalization

In this subsection we will argue that the "peculiar" phenomenon of double determination in Swedish can further be naturally accounted for as a result of contact-induced

grammaticalization. More precisely, following Dahl (2004b) (see also Heine and Kuteva 2006; Kuteva forthcoming), we will show that the Scandinavian dialectal varieties with both a preposed and a postposed definite article in the adjective-noun phrase are located in an area where two opposite grammaticalization patterns meet. We will show that the resolution of the potential conflict arising from the convergence of the two competing structures is the creation of a Buffer Zone, i.e., a zone where an areal overlap leads to a merger of the two patterns (for the notion of Buffer Zone, see Stilo 1987, 2005). The result of this merger is the double marking of definiteness as we find it in Swedish (and in most dialects of Norwegian).

Let us recall the facts about the grammatical distribution of the two kinds of definite article in Scandinavian again. In a simplified format, Danish (marginally also Norwegian) uses suffixed articles in noun phrases which do not contain a preposed modifier but preposed articles when the noun phrase contains an adjective or a quantifier—that is, the two kinds of articles occur in complementary distribution. In Swedish (and most forms of Norwegian), on the other hand, there is typically both a preposed and a suffixed article in noun phrases containing preposed attributes. In other words, the preposed and the suffixed articles **do not** co-occur in Danish, whereas in Swedish (and in most dialects of Norwegian), the preposed article **does not** occur without the suffixed article.

Now, the core of the explanation of the "double determination" phenomenon in terms of contact-induced grammaticalization is the following fact: the extent to which the preposed and the postposed definite articles are grammaticalized differs from one language variety to another, and these differences exhibit a clear *areal patterning*. The detailed description by Dahl (2004b) reveals that—with the exception of Icelandic and Faroese—spoken Scandinavian varieties constitute a dialect continuum cutting across national boundaries. With reference to definite marking in adjective–noun combinations, there is a clear areal distribution on the basis of relative degree of grammaticalization. Dahl observes that there are two separate grammaticalization processes giving rise, respectively, to a suffixed- and a preposed-article area: "In this perspective, it is natural to see 'double determination' as one possible outcome of the competition between two different grammaticalizing definite articles" (Dahl 2004b: 178).

What a brief diachronic survey reveals is that the suffixed article is the historically older one and goes back to Old Scandinavian. With the agricultural expansion from 1050 to 1350, Old Scandinavian spread out to the North. As there is not so much fluctuation of population in the North, traits of the old forms remained preserved and are to be found as conservative patterns in the dialects of northern Sweden and Swedish-speaking Finland and Estonia. Thus the suffixed-article use is the older one, having reached its fullest development in northeastern Scandinavia in the "peripheral Swedish dialect area". Here this use has been expanded to an extent which we do not see anywhere else in Europe, so that the article can even be used in contexts involving non-specific indefinite reference (see Dahl's 2004b: 174 discussion of "low referentiality uses"), as in (13).

(13) Nederkalix, Norrbotten (Dahl 2004b: 174)

Jä skå tåla åom för dä, måmme, åt jä ållti veillt hå i kjaatt
I shall speak.INF about for you.DAT mother that I always want.SUP have a cat
män hä gja jo ät håå ***kjatta*** *når man båo ini i*
but it go.PRES as.you.know not have.INF cat.DEF when one live.PRES in a
höreshöus.
rent-house
'I want to tell you, Mother, that I have always wanted to have a cat—but it isn't possible to have a cat (lit. **the cat**) when you live in an apartment house'

Note that not only in languages like English—which constitutes an example of a highly grammaticalized preposed article—but also in Bulgarian, a language with a well-established postposed article, the counterpart of the non-specific indefinite referent cannot be rendered by means of a definite article, as in (14).

(14) Bulgarian

Iskam da ti kaža, majko, če vinagi săm iskala da imam kotka– samo
want.1.SG.PRES to you tell mother that always am wanted to have cat only
če ne e văzmožno da imaš ***kotka****, kogato živeeš v*
that not is possible to have.2SG.PRES cat when live.2SG.PRES in
blok
apartment.house
'I want to tell you, Mother, that I have always wanted to have a cat—but it isn't possible to have a cat (lit. **cat**) when you live in an apartment house'

In fact, the expanded use of the suffixed definite article in the peripheral Swedish dialect area seems to be a precursor of what Hawkins (2004: ch. 4) identifies as the last stage of development of definite articles where the article is recruited for purely syntactic purposes, so that, in the end, all connections to definiteness/indefiniteness are lost—as, for instance, in Tongan and Maori, Polynesian (Hawkins 2004: 85).

At the other extreme there is Denmark, displaying the most restricted use of the suffixed article; central and southern Sweden and Norway form an intermediate zone.

Conversely, the preposed article is the innovation in the Scandinavian area and exhibits the Western European pattern of preposed-article use. Its occurrence is strongest in Danish, which also is geographically closest to the western article-languages. With the Danish colonization of southern Sweden, the Danish preposed article pattern came into use. Thus, preposed article-use is much higher in southern Sweden, higher than in Standard Swedish, whereas the preposed article is used below average in northern Sweden. Dahl (2004b: 176) divides the Continental Scandinavian dialect continuum into "five different sub-areas, with respect to the strength of the preposed article, ordered in decreasing strength from the south-west towards the northeast as follows: (i) SW Jütland (where the **preposed article** is in general use as a definite article); (ii) the rest of Denmark; (iii) Norway and southern and western Sweden; (iv) central Sweden; (v) northern Sweden".

The result of this areal distribution is that there is an intersection zone in the center, where the two article areas overlap—that is, a Buffer Zone—so that there is "double determination", while at the southern borderline there appears to be a gradual transition towards a canonical situation in Standard Average European, where there are only preposed but no suffixed articles.

In sum, the two distinct grammaticalization processes described above have different centers of gravity—the preposed article in the south, and the suffixed article in the north—and it is precisely in the area of high-intensity contact, namely in the geographically transitional zone, that these two grammaticalization patterns overlap, giving rise to double marking.

It remains largely unclear what exactly the historical processes were that contributed to this areal distribution in the Scandinavian dialect continuum. What is beyond reasonable doubt is that this distribution correlates with differences in grammaticalization on the one hand and areal gradience on the other, in that there is both a south-to-northeast cline and a northeast-to-south cline with regard to relative degrees of grammaticalization. In view of these correlations there is reason to hypothesize that language contact and grammaticalization must jointly have played some role in the diffusion of the two types of definite articles in Continental Scandinavia.

Note that the "double determination" phenomenon in Swedish constitutes a typical Buffer Zone behavior observed in other regions of the world, too. Thus Stilo (forthcoming) describes a Buffer Zone in what he identifies as the Araxes Sprachbund area—an area encompassing a number of Iranian, Armenian, Turkic, Semitic, Kartvelian, and Lezgic language varieties—where two of a number of isoglosses meet (see also Kuteva forthcoming). One of these isoglosses involves oblique pronominal enclitics functioning as possessives with nouns, that is, a postposed possessive, and the other isogloss involves independent pronominal forms functioning as possessives with nouns, that is, a preposed possessive. Now, what happens in the area where these two isoglosses meet, i.e., in the Buffer Zone? Stilo clearly shows that the Buffer Zone has both (i) alternation between the preposed and the postposed (enclitic) forms, and (ii) circumposed (i.e., doubly marked) possessive pronominal expressions. Armenian is one of the languages which belong to the Buffer Zone, and it clearly exhibits a circumposed possessive, as in (15).

(15) Armenian (Stilo forthcoming)
im tun=əs
my house=my
'my house'

As becomes clear from (15), Stilo's circumposed possessive pronouns can readily be translated into what we have been referring to as "double determination", or simply doubly marked expressions.

At this stage of research, Stilo's work is centered on the geographical distribution of synchronic facts; grammaticalization phenomena remain outside his scope of investigation. At a later stage, however, it might turn out to be a feasible exercise—both

in contact-induced grammaticalization and in Sprachbund linguistics—to check if the above Buffer Zone in the Araxes Sprachbund exhibits a similar kind of opposing and gradual grammaticalizing clines meeting in space and time as in the Buffer Zone in Scandinavian. The fact that the possessive markers in the Buffer Zone languages of the Araxes Sprachbund can be historically traced back—at least in the case of Armenian—to another category, the demonstrative, makes such an exercise a highly promising endeavor.

9.4 CONCLUSIONS

Some of the most fundamental questions facing grammaticalization theory are: just how strong is the explanatory power of grammaticalization? Does grammaticalization explain the regular aspects of grammar, or is it only helpful in accounting for the irregular, the exceptional facts?

In our previous work (Heine et al. 1991; Kuteva 1998, 2001; Heine and Kuteva 2001, 2002; Heine 2003) we have shown the explanatory potential that grammaticalization has in explaining the sizeable amount of universal grammaticalization developments that have been identified so far as regular, cross-linguistically persistent paths of historical change. At the core of this explanation—we have argued—are universal conceptualization principles such as (i) making use of concrete, easily accessible and/or clearly delineated notions in order to build less concrete, less easily accessible and less clearly delineated meaning contents, and (ii) conceptual transfers such as metaphor, metonymy, etc. The reader is also referred to the work of Bybee et al. (1994), Bybee (2006b), Bybee (this volume), among others, on the mechanisms of change—which are universal and relatively few in number—as the explanation of cross-linguistic universals.

In this chapter, it has been our goal to show that grammaticalization theory can explain not only the regular patterns, the non-exceptions, but also the exceptions. For both of these, it is not only language-internal grammaticalization but also contact-induced grammaticalization that is important.

To claim that a particular language-contact situation must have played at least an accelerating and/or reinforcing role for a particular language-change process is always a risky enterprise. The reason is that it is not easy to tease apart all the factors involved in language-change processes. In some cases, however, a cluster of observations may come to point in the same direction, indicative of language contact working in conspiracy with language-internal change (on general "test conditions" that indicate whether a pattern arose from contact-induced grammaticalization or language-internal grammaticalization only, cf. Heine and Kuteva 2006: chs. 1 and 8).

In the present paper, we discussed two—at first sight—exceptional situations relating to the area of definiteness marking in European languages. Employing the theoretical apparatus of what we have called elsewhere (Heine and Kuteva 2006; Kuteva

and Heine forthcoming) *integrative* grammaticalization theory—that is, a theory of grammaticalization taking into account both contact-related and non-contact-related situations—we proposed an explanation of these seeming "exceptions" in terms of (i) a *diachronic overlap stage* in the course of language-internal grammaticalization and (ii) both a *diachronic overlap stage* and a *synchronic Buffer Zone* in the case of contact-induced grammaticalization. In the latter case, the two explanations reinforce each other, and are fully in accordance with the insight gained from the study of areal grammaticalization: contact-induced grammaticalization simultaneously implies both external and internal language change.

Theologians say: "Language change is a consequence of man's arrogance (as manifested in the Tower of Babel)". How about claiming one can explain each and every exception in the development of each and every human language? Wouldn't *that* be arrogance?

Here we are not claiming we are in a position to explain each and every exceptional structure or development in grammar. For example, it appears that there exists at least one group of language varieties—Kæbætei and Kelasi (two Central Tati dialects, 12 kilometers apart, mutually intelligible but distinct, with no intervening villages), with thanks to Don Stilo (fieldwork)—where the postposition *æmra* having initially a comitative meaning (historically deriving from **hæm-ra* 'same-road/sharing the same path') developed an instrumental function (a development that is well attested in grammaticalization studies) which later on, however, came to function as an ablative marker.[2] Now, from a grammaticalization point of view, this is an exception, something which we would not predict since it is not an otherwise well-attested grammaticalization pathway. It may, of course, very well turn out that, at a closer look, this attestation lacks important details/stages in the gradual step-by-step development of the structure, which—if known—would give us a clue as to how the ablative meaning could have arisen out of a form which at an earlier stage of its development had a comitative/instrumental function. Note that this is the story of a number of seeming exceptions in the behavior of linguistic structures. As an illustration, one can point out the—at first sight, odd—development of the future tense marker in Hup (a Vaupés-Japura language, spoken in Brazil/Colombia) from a noun meaning 'stick/long wooden object' (Epps 2007). A number of grammaticalization paths resulting into a future grammatical morpheme have been attested in the languages of the world (Heine and Kuteva 2002), but neither of them involves a noun with the above meaning. However, if one makes a detailed study of the entire range of uses of the Hup form, it turns out that one of those involves the use of the noun as an *instrument for a certain purpose*, and on the basis of such a detailed synchronic study, Epps builds a convincing case for the grammaticalization of the nominal form into a futurity-denoting verbal form via an intermediate stage involving a purposive meaning. Now,

[2] There are at least thirteen more language varieties—a number of them spoken in the Araxes Sprachbund—where one comes across a synchronic comitative-(and-instrumental)-and-ablative polysemy with adpositions (Don Stilo, p.c.).

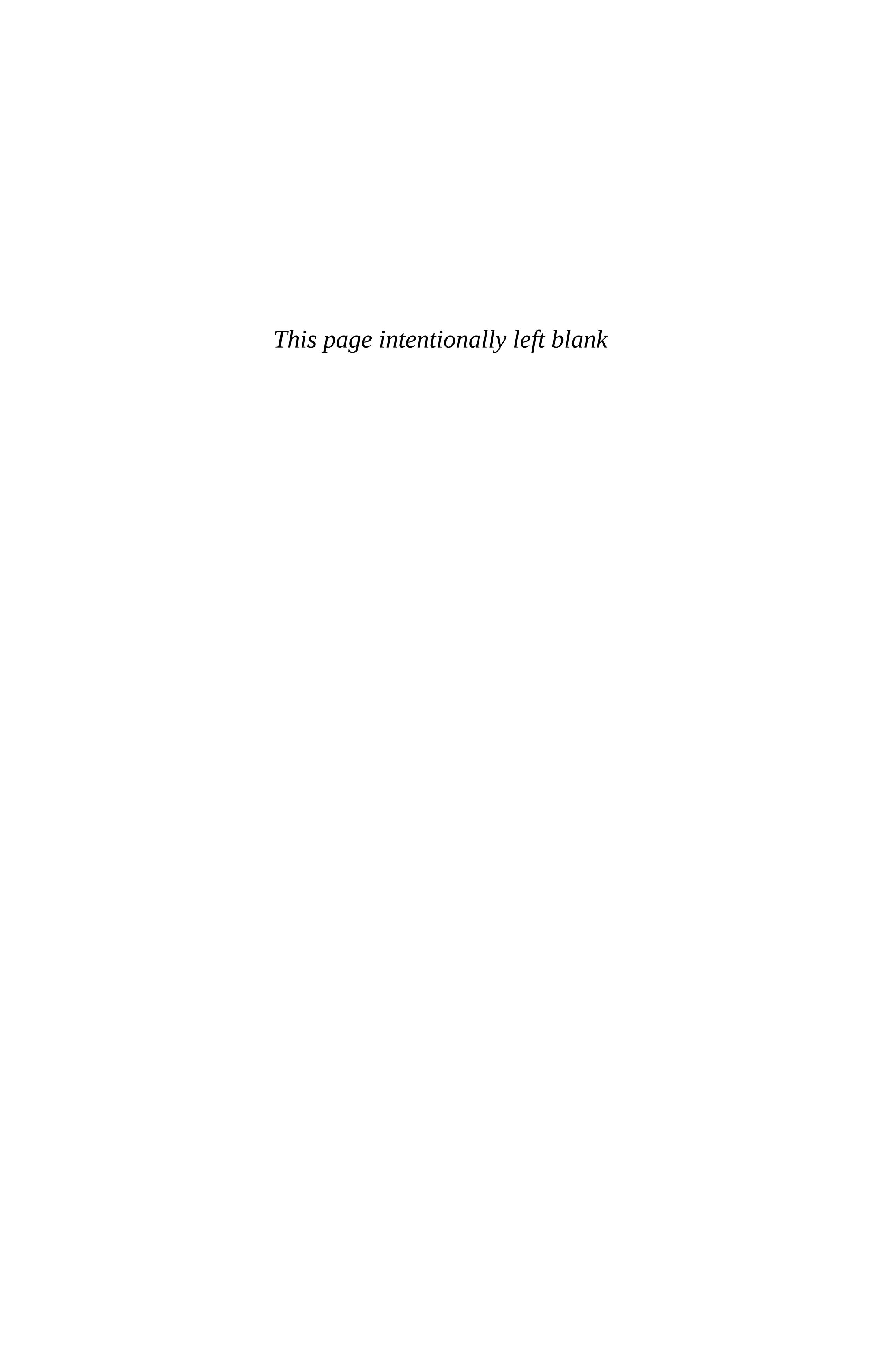
a purposive meaning is among the meanings which are most frequently attested as extending to future meaning. Thus the Hup example cannot sustain a status of an exceptional language with respect to its future-tense marker.

Yet, the fact remains that in the absence of any hard-and-fast evidence about the step-by-step development of the postposition *æmra* in Kæbætei and Kelasi, we can only say that we are not in a position to reconstruct the process, and thus explain the exceptional pattern of behavior.

Beside the cases specific to a particular language or to a particular culture, however, there are legions of cases which reflect general aspects of human cognition, and we claim that these aspects are manifested in linguistic structure, irrespective of the fact that in every individual language there are also additional factors pertaining to the entire system of that language and the relationships within it. These general cognitive aspects show through the development of language grammar, no matter whether we are dealing with a single language grammar developing in relative isolation from other grammars or with the grammars of more than one language which—through intense social and geographical integration resulting in mutual diffusion—have come to develop their grammars "together". In other words, here we have tried to articulate a more flexible approach to the study of linguistic change—that is, the one adopted in recent grammaticalization studies. That grammaticalization has a high explanatory value with regard to what are otherwise regarded as "significant distortions of the underlying syntactic structure" is also to be seen in the fact that recently there have appeared attempts (Zoe Wu 2004) to apply it as an operational tool in the study of language-change phenomena also in the Minimalist/Chomskyan approach to language. What the results of such attempts will be is something we will be able to critically assess only in the future.

PART V

Phrase Structure: Modeling the Development of Syntactic Constructions



This page intentionally left blank

10

The Classification of Constituent Order Generalizations and Diachronic Explanation

John Whitman
Cornell University

10.1 INTRODUCTION

Greenberg's (1963) constituent order generalizations have long been a battleground for opposing modes of linguistic explanation. Diachronic explanations for some of the generalizations begin with Givón's pioneering work (1975, 1979) and are the focus of Aristar (1991). The much larger catalogue of proposed explanations based on synchronic considerations—including constraints imposed by language processing or discourse function—range from Hawkins' (e.g., 1983) processing-based accounts to much work in the generative tradition, most recently Kayne (1994) and research inspired by Kayne's antisymmetry program. Apart from etiology, the Greenbergian[1] generalizations have also been subject to various classificatory schemes, usually based on internal properties such as complexity (implicational versus unconditional universals) or strength (statistical versus absolute universals). This paper proposes a different classification of the Greenbergian constituent order universals, which extends naturally to larger compendia (e.g., Dryer 1992; Plank 2003; Haspelmath et al. 2005). Based on this classification, I suggest that the best-known generalizations of this type, cross-categorial universals, arise most plausibly through language change (and thus are usually statistical). Two other types, which I label hierarchical and derivational universals, are true candidates for principles of synchronic grammar.

[1] The descriptor is due to Dryer (1992). I intend it here to apply both Greenberg's own generalizations about constituent order and similar generalizations by others.

10.2 THE CLASSIFICATION

Setting aside their logical or evidential properties to examine Greenberg-style constituent order generalizations in terms of the kinds of linguistic data they cover, I propose to distinguish three subtypes.

(1) **Constituent order generalizations (universals)**
  a. **Cross-categorial generalizations** reference the internal properties of two or more categories irrespective of their relationship in a particular structure.
  b. **Hierarchical generalizations** describe the relative position of two or more categories in a single structure.
  c. **Derivational generalizations** describe the relative position of two or more categories at the end of a syntactic derivation.

Cross-categorial generalizations are the best known in the Greenbergian inventory. They are the main topic of Hawkins' 1983 monograph and subsequent work, and the specific focus of constructs in generative grammar such as the Head Parameter (Chomsky 1981). In much typological work, the significant and interesting Greenbergian generalizations are held to be precisely the cross-categorial generalizations (what Dryer 1992 calls the Greenbergian correlations). I will argue that this focus on cross-categorial generalizations, in formal and functional work alike, has led to a certain skewing of linguists' expectations about what a prototypical syntactic universal should look like: for example, that it should be statistical.

Greenberg's Universal 3 is an example of a **cross-categorial** generalization:

Universal 3. Languages with dominant VSO order are always prepositional.

(Greenberg 1963: 78)

Universal 3 correlates constituent order internal to one category (S; that is, the clause) with constituent order internal to another (PP; that is, adpositional phrases). Such generalizations hold regardless of the structural relationship between the two categories; for example, instances of category S may or may not contain PP, but this relationship is irrelevant to the interpretation of Universal 3.

Hierarchical generalizations refer to the relative position of two categories within a single syntactic structure. Greenberg's Universals 1 and 14 are hierarchical universals:

Universal 1. In declarative sentences with nominal subject and object, the dominant order is always one in which the subject precedes the object.

Universal 14. In conditional statements, the conditional clause precedes the conclusion as the normal order in all languages. (Greenberg 1963: 78)

Greenberg formulated Universals 1 and 14 to specify the relative order of pairs of categories (subject, object; conditional, consequent) in a single structure. Greenberg's formulations reference linear order, not structural position, but the categories he refers to stand in a fixed hierarchical relation at some level of representation in many

structurally oriented theories of grammar. Two notions are relevant here. The first is the idea that there is some relationship between the structural positions of two constituents and their linear order. This idea is controversial, but there are some broad points of relative consensus. For example, Universal 1 follows, at an appropriate level of representation, if (i) subjects originate in the specifier of a projection that contains the object, and (ii) specifiers always precede their heads, as in (2).

(2) $[_S$ Specifiers $[_{VP}$ precede heads and complements]]

Likewise, the conditional clause can be held to occupy a higher structural position than the consequent in conditional statements, again at an appropriate level of representation, if conditionals are generated in the specifier of a projection that contains the consequent clause, as in (3)

(3) $[_{S'}$ If conditionals are specifiers of S' $[_S$ they precede the consequent]]

The second key notion is the idea of appropriate level of representation. Greenberg's interest, of course, was in surface constituent order, a tradition maintained in most typological research. Neither Universal 1 nor 14 holds as an absolute universal about surface order across languages. Universal 1, for example, has been known to be statistical, not absolute, as a characterization of surface constituent order, since the first studies of object-initial languages (Derbyshire 1977). I will argue, however, that these universals are absolute when applied at an appropriate level of representation.

I have defined hierarchical and derivational universals in (1b–c) in such a way that the latter are actually a subcase of the former. In practice I will restrict hierarchical universals to cases where underlying constituent order is key to explaining Greenberg's generalization.

Derivational generalizations also involve ordering relations between categories in a single syntactic domain, but they refer to relative position at the end of the syntactic derivation. This classification is unavoidably theory-internal: it requires a theory where surface word order is syntactically derived. Let us take Greenberg's Universal 7 as example.

Universal *7*. If in a language with dominant SOV order, there is no alternative basic order, or only OSV as the alternative, then all adverbial modifiers of the verb likewise precede the verb. (This is the rigid subtype of III.) (Greenberg 1963: 80)

Universal 7 is a derivational generalization within a framework such as Kayne's (1994). On Kayne's approach, underlying word order is universally SVO. OV order can be derived in two ways: by fronting the object (and perhaps other core grammatical arguments), or by fronting the VP, minus the verb. The first option derives the S-O-V-other order found, for example, in Mande (Gensler 1994), as in (4). The second option derives Greenberg's "rigid subtype" of SOV (Type III), as in (5).

(4) S XP O V t_O YP

(5) S [XP O t_V YP ... $]_{VP}$ V t_{VP}

Since the option in (5) derives OV order by fronting the VP around the verb, it necessarily fronts all constituents within VP, including adverbs, as well. Universal 7 can be interpreted as a derivational universal because the universal crucially references constituent order after the pattern in (5) has been derived.

Note that Universal 7 has the form of an implicational universal, just like Universal 3. But Universal 7 refers to the relative ordering of elements within a single clause, while Universal 3 refers to the position of elements in distinct domains (V in S, P in PP). In this paper I argue that derivational generalizations, like hierarchical generalizations, are candidates for true universals.

In the following section I briefly discuss the statistical status of the Greenbergian generalizations. Section 10.4 examines the status of the apparently strongest cross-categorial generalizations. Sections 10.5 and 10.6 examine derivational and hierarchical generalizations respectively.

10.3 STATISTICS

The status of "statistical" versus "absolute" universals is widely discussed in the typological literature, with some linguists arguing (e.g., Dryer 1998) that all meaningful universals are statistical. As pointed out by Aristar, non-absolute generalizations are poor candidates for synchronic explanations based on universal properties of grammar or cognition (1991: 4). I believe that Dryer is right in asserting that cross-categorial constituent order generalizations are inherently statistical. However, I will suggest that in this regard they are quite distinct from the other two types of universal distinguished above.

Let us begin by examining the veridical status of Greenberg's own generalizations. The universals that I classify as cross-categorial are listed in Table 10.1 with the statistical status assigned to them by Greenberg.

Fourteen of Greenberg's twenty-five exclusively syntactic generalizations are cross-categorial. Eight of these are characterized by him as exceptionless: 3, 5, 12, 13, 15, 16, 21, and 24. Only two of the twenty-five syntactic generalizations are hierarchical, 1 and 14, discussed in section 10.2. According to Greenberg, the first of these is statistical, the second absolute. As noted above, the apparent paucity of hierarchical universals is directly related to the role of derivational universals: derived orders obscure hierarchical generalizations. In section 10.5, I classify seven of Greenberg's twenty-five syntactic generalizations as derivational (6, 7, 10, 11, 19, 20, and 25); all of these are exceptionless. In section 10.4, I show that the difference in strength between cross-categorial and derivational generalization is even sharper than Greenberg's paper indicates. Subsequent research shows that all of the eight cross-categorial universals classified by Greenberg as exceptionless are in fact statistical.

TABLE 10.1. Greenberg's cross-categorial generalizations

No.	Universal (Greenberg 1963)	Strength
2	In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes.	almost always
3	Languages with dominant VSO order are always prepositional.	always
4	With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional.	overwhelmingly greater than chance
5	If a language has dominant SOV order and the genitive follows the governing noun, then the adjective likewise follows the noun.	always
9	With well more than chance frequency, when question particles or affixes are specified in position by reference to the sentence as a whole, if initial, such elements are found in prepositional languages, and, if final, in postpositional.	well more than chance
12	If a language has dominant order VSO in declarative sentences, it always puts interrogative words or phrases first in interrogative word questions; if it has dominant order SOV in declarative sentences, there is never such a rule.	always
13	If the nominal object always precedes the verb, then verb forms subordinate to the main verb also precede it.	always
15	In expressions of volition and purpose, a subordinate verbal form always follows the main verb as the normal order except in those languages in which the nominal object always precedes the verb.	always
16	In languages with dominant order VSO, an inflected auxiliary always precedes the main verb. In languages with dominant order SOV, an inflected auxiliary always follows the main verb.	always
17	With overwhelmingly more than chance frequency, languages with dominant order VSO have the adjective after the noun.	overwhelmingly more than chance
18	When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency, do likewise.	overwhelmingly more than chance
21	If some or all adverbs follow the adjective they modify, then the language is one in which the qualifying adjective follows the noun and the verb precedes its nominal object as the dominant order.	always
22	If in comparisons of superiority the only order, or one of the alternative orders, is standard-marker-adjective, then the language is postpositional. With overwhelmingly more than chance frequency if the only order is adjective-marker-standard, the language is prepositional.	always (part 1); overwhelmingly greater than chance frequency (part 2)
24	If the relative expression precedes the noun either as the only construction or as an alternate construction, either the language is postpositional, or the adjective precedes the noun or both.	always

Similar results are provided by a larger sample of proposed language universals. The Konstanz Universals Archive (Plank 2003) turns up 430 records on a keyword search for "order"; of these some 269 are generalizations about syntactic constituent order. Some 154 of the constituent order generalizations are cross-categorial; of these, 106, just over two-thirds, are identified by the archive compilers as statistical. Of the remaining 48 constituent order generalizations classified as "absolute", a substantial number have counter-examples noted by the database compilers or previous researchers, indicating that they are, in fact, statistical, although originally proposed by their authors as absolute.

10.4 CROSS-CATEGORIAL GENERALIZATIONS

I suggested in the previous section that all cross-categorial constituent order generalizations are statistical. As noted there, virtually all of the eight cross-categorial generalizations classified by Greenberg (1963) as exceptionless have since been found to be statistical.

Greenberg himself (1963: 107) notes Papago (Tohono O'odham) as an exception to Universal 3; further counter-examples are provided by Payne (1986), while Dryer in Haspelmath et al. (2005) lists six languages which are postpositional with dominant VSO order (out of a total of thirty-eight postpositional languages with dominant VO order).[2]

Universal 5 is a nested conditional: the population of languages that are OV and that have noun–genitive order is very small to begin with (Dryer 1992). Even so, Plank (2003) attributes to Dryer the observation that Tigre counter-exemplifies Universal 5. Tigre has OV, noun–genitive, and adjective–noun order, as shown in (6).

(6) Tigre NP internal order (Raz 1983: 95)

a. Galab ʾət ʾətyopya latətrakkab [hatte nəʾis dəgge] ta.
Galab in Ethiopia which.is.found one small town is
'Galab is a small town which is found in Ethiopia.'

b. ʾAze hatte məʾəl ʾət [hədday ʾad waʾaga] fararaw.
Now one day to wedding.party family guenon they.went.out
'One day they went out to the wedding party of the family of the guenon.'

Universal 12 is a cross-categorial universal because it correlates the word-order properties of two distinct categories (in this case, sentence types): interrogative and declarative clauses. It is statistical. Dryer in Haspelmath et al. (2005) lists 16 VSO languages which do not place *wh*-phrases in initial position in content questions, as

[2] Relevant chapters of Haspelmath et al. used in this paper include Dryer (2005a, 2005b, 2005c, 2005d).

opposed to 42 which do. Some 52 SOV languages have initial *wh*; 225 SOV languages do not.[3]

Counter-examples to Universal 13 have not been widely discussed, but they occur in Tibeto-Burman languages classified by Greenberg as the "rigid subtype" of type III (SOV), such as Burmese. The relevant pattern involves "fore-and-aft concatenations" (Matisoff 1973: 248), where a subordinate verb appears on either side of the higher verb, depending on its meaning. For example, Burmese (SOV) *pè* 'give' follows the subordinate verb when it is interpreted as a benefactive (the pattern predicted by Universal 13), but precedes when interpreted as a permissive causative.

(7) Burmese *pè* benefactive / causative
   a. Canáw kà maunlèi ko màun pè-dɛ.
      I NOM boy DAT / ACC drive give-REAL
      'I drove for the boy, I did the boy the favor of driving.'
      (modified from Okell and Allot 2001: 120)
   b. Canáw kà maunlèi ko pè màun-dɛ.
      I NOM boy DAT / ACC give drive-REAL
      'I let the boy drive.'

Along these same lines, exceptions to Universal 15 are cited by Dryer (1992: 94), and to 16, 21 and 22 by Pickett (1983)[4] (the authors of these counter-examples are noted in Plank 2003).

Suppose, then, that it can be established that all cross-categorial constituent order generalizations are statistical. What does this say about the status of these generalizations in grammar? There are two views about this. Since the 1980s, one prominent view has held that cross-categorial correlations (to use Dryer's term) are the reflex of a parameter of Universal Grammar, the Head Parameter. This approach would explain Universal 4, for example, by claiming that children set a single value [final] for the categories VP and PP in acquiring a language such as Burmese.

(8) Burmese VP, PP
   a. [kà màun-tɛ.]$_{VP}$
      car drive-REAL
      'drive a car'
   b. [Yankon ko]$_{PP}$ θwà-tɛ.
      Rangoon to go-REAL
      '(go) to Rangoon'

[3] Figures like these were obtained using the extremely useful interactive CD accompanying Haspelmath et al. (2005), which allows searches for combined features such as constituent order and position of *wh*-phrase.

[4] Pickett (1983: 540) cites Yaqui, Mixe, and Zoque as counter-examples to the second part of Universal 16: in these languages, at least some varieties of which she classifies as Greenberg's type III (OV, postpositional), the inflected verb 'be able' precedes the main verb. She cites Mixe and Seri as counter-examples to Universal 21 (1983: 544): these languages have adjective–adverb and noun–adjective but OV order. She cites Tepehua (1983: 542) as a counter-example to Universal 22: this language has standard–marker–adjective order in comparatives, but it is prepositional.

The problems with this approach are well known, however. First, recall that the relevant generalizations are statistical, so there are exceptions. Persian is OV but prepositional:

(9) Persian (OV) prepositions (Windfuhr 1987: 534)
[be mán]$_{PP}$ dād
to me gave
'[Someone] gave it to me.'

Other languages have mixed values for the closed-class category P: Amharic and Chinese, for example, have both prepositions and postpositions.

(10) Chinese prepositions and postpositions
a. Tāmen [cóng Měiguó]$_{PP}$ lái.
3PL from U.S. come
'They come from the U.S.'
b. [Zhuōzi shàng]$_{PP}$ yǒu yī tái pòsùi de diànnǎo
table on is one CL broke COMP computer
'On the table is a broken computer.'

Crucially, there is no evidence that "exceptional" languages such as Persian or "mixed" languages such as Chinese are more difficult to acquire, or that delays occur in the acquisition of the exceptional or mixed category types. Evidence from first-language acquisition suggests that acquisition of basic word order is very early, regardless of the word-order properties of the target language (Wexler 1998). These facts have contributed to a rejection of the Head Parameter as a component of Universal Grammar (Kayne 1994; Newmeyer 2005).

The second approach to explaining cross-categorial constituent order generalizations appeals to diachronic processes. Givón (1975: 82), for example, points out that adpositítions are frequently derived from serial verbs. In VO languages, this results in prepositions, while in OV languages it results in postpositions. Adpositions also develop from relational nouns; thus the "mixed" property of Chinese PPs is the historical consequence of prepositions derived from verbs such as *cóng* 'from' in (10a) (Chinese is VO) and postpositions derived from relational nouns such as *shàng* 'on' in (10b) (Chinese is modifier-noun).[5] Complementizers and auxiliaries are well known to have similar sources; a very simplified sketch of this kind of development is given in Table 10.2, with examples, mostly well known, from English.

This diachronic mode of explanation predicts that PP and VP internal order will be consistent in a language to the extent that adpositions have a diachronic source from verbs. Mixed orders such as Chinese (10) result when VP and NP vary in head–complement order and Ps derive from both. The source of Persian [P NP] order is complex, but the Modern Persian preposition *be* results from reanalysis of a preverb

[5] See Whitman and Paul (2005) for a description of the derivation of the preposition *cóng* 'from' from a verb meaning 'follow, accompany'.

TABLE 10.2. "Grammaticalization" type cross-categorial shifts

Category	Source	Example
Comp	Verb	English *have* > <of> in *I should of gone* (Kayne 1997)
Comp	Adposition	English *for* + infinitive
Aux	Main verb	English *will*
P	Verb	English *concerning, regarding*
P	Relational Noun	(10b)

and an earlier preposition (Lazard 1986), which takes over the functions of the Middle Persian preposition *pa(d)* < Old Persian *patiy*.[6] But whatever the historical source of the "mixed" or "exceptional" cross-categorial pattern, this diachronic information is nowhere in the synchronic grammar, and is irrelevant to the child acquiring Persian or Chinese. First-language learners are able to acquire any of the three patterns, unaided by parametric information of the X-bar theoretic type.

Aristar (1991) objects that this mode of explanation may work for the kinds of correlations in Table 10.2, where there is a clear diachronic "path" from a lexical category (verb, noun) to a closed-class category (adposition, complementizer). But what about correlations across the lexical categories N and V? A common position in the generative literature is that relative head–complement order tends to hold across NP and VP as well, yielding "consistent" values for the Head Parameter in languages that are "strictly head final" (Japanese) or "strictly head initial" (Tagalog). Accepting this view, Aristar attempts to construct a diachronic scenario for explaining word-order correlations across NP and VP.

But there is good evidence that this attempt is mistaken. Dryer (1992), using a far larger sample of languages than Greenberg, shows that noun–genitive and noun–relative order does not correlate with verb–complement order, while P–complement order does. This result is extremely significant for the argument that cross-categorial word-order generalizations have a diachronic source: cross-categorial generalizations about constituent order hold in just the case where grammaticalization is abundantly attested (e.g., verb, adposition). Where paths of grammaticalization are rarer or arguably nonexistent (noun phrase, verb phrase), cross-categorial generalizations do not stand up. Dryer's results further strengthen the view that significant cross-categorial word-order generalizations arise through language change.

It remains the case that the veridicality of some cross-categorial generalizations is very high; some appear to have only small numbers of counter-examples. Thus as we have seen, Plank (2003) cites no counter-examples to Universal 13, although we were able to cite the Burmese pattern in (7) and Matisoff's (1973) more general discussion of such patterns in Tibeto-Burman. In such cases, we must imagine that the diachronic combination of circumstances that give rise to the counter-exemplificatory pattern is

[6] I am indebted to Michael Weiss for explaining the Persian diachronic facts and introducing me to Lazard (1986).

relatively rare.[7] For example, patterns such as (7) with order O–main verb–subordinate verb in strict OV languages seem to be restricted to serializing OV languages (see Carstens 2002 for similar examples from Ijọ). Both the benefactive pattern in (7a) and the causative pattern in (7b) are diachronically derivable from coordinate structures; the order of "give" and the accompanying verb is determined by the temporal order of events in the source coordinate structure, as in (11). (See Newman 1996 for the development from permissives to causative.)[8]

(11) a. buy the food and give the boy pro_{food} > buy give (benefactive)
b. give the boy the food and pro_{boy} eat pro_{food} > give eat (> permissive)

On this analysis, the two complementation patterns in (11), both attested in Burmese (7) with the same verb 'give', are simply the result of the two different diachronic sources for the pattern, both from coordinate structures. When 'give' comes second in the coordinate source structure, grammaticalization yields a benefactive verb. When 'give' comes first, grammaticalization yields a permissive. Both patterns preserve the basic argument structure of the source coordinate structures: in the benefactive (11a), the subject of 'give' is also the subject of 'buy'; in the permissive (11b), the goal of 'give' is the subject of 'eat'. The typological rarity of the pattern reflects two factors: the relative rarity of strict OV serializing languages (Schiller 1990: 396) and the analogical pressure imposed by coexisting verb-causative patterns in the same languages (see, e.g., Okell and Allott 2001 for description of the Burmese postverbal causative).

10.5 DERIVATIONAL GENERALIZATIONS

Following the same strategy as in the preceding section, in this section I examine the derivational universals 6, 7, 10, 11, 19, 20, and 25, all characterized by Greenberg as exceptionless.

Universal 6. All languages with dominant VSO order have SVO as an alternative or as the only alternative order. (Greenberg 1963: 79)

This universal has apparently remained unchallenged in the literature subsequent to Greenberg. The derivational relationship between VSO and SVO is directly accounted for on the verb-raising analysis developed in the transformational literature (Emonds 1980; Sproat 1985; McCloskey 1991). Under this approach, VSO order is derived by raising the verb over the subject out of a constituent VP. SVO order is predicted to

[7] The insight that typological "oddities"—exceptionally rare syntactic or morphological patterns—may result from combinations of historical developments (which themselves may be relatively common) is due to Alice Harris (this volume).

[8] Newman (1996) suggests that a purposive structure such as 'buy the book to give to the boy' may be an intermediate stage in the development of 'give' to a permissive causative. But Tibeto-Burman gives no evidence for postverbal purposive structures.

occur in contexts where verb raising is blocked, such as non-finite clauses. In fact, from a derivational standpoint, the occurrence of SVO order is dependent on the existence of such contexts: if a VSO language lacks non-finite verbal constructions, we might expect it not to attest SVO order. On this interpretation, a refined version of Universal 6 might be as in (12).

(12) All languages with dominant VSO order have SVO as an alternative order if they have non-finite structures; this is the only alternative order involving verb placement.

The label "derivational generalization" might suggest that generalizations of this type are of interest only to transformational theories, and in the remainder of this section I will focus on transformational accounts. But in principle, any theory which maps between syntactically relevant levels of representation may capture the generalizations that I have called "derivational". Thus the account of VSO order in languages such as Welsh in Lexical-Functional Grammar (Sadler 1997; Bresnan 2001: 127–131) relies on the mapping between f-structure (where the lexical verb is the predicate of the clause) and c-structure (where the finite lexical verb occupies the position of I(NFL), the head of the finite clause above the subject NP). This account too makes finiteness the crucial feature for predicting the occurrence of SVO order in predominantly VSO languages.

In section 10.2 I described how Universal 7, repeated below, is captured in Kayne's (1994) transformational account of constituent order variation.

Universal 7. If in a language with dominant SOV order, there is no alternative basic order, or only OSV as the alternative, then all adverbial modifiers of the verb likewise precede the verb. (This is the rigid subtype of III.) (Greenberg 1963: 80)

However, Kayne's is not the only derivational account of Universal 7. Any analysis which shares the insight that units of clausal structure above the verb phrase, such as tense and subordination/clause type markers (complementizers), are present in Greenberg's "rigid subtype" of type III (strictly head-final) languages must posit a mechanism to account for the fact that the verb and these higher units of structure form an inseparable sequence at the surface level of representation. Under such treatments, the syntactically relevant representation for a tensed matrix clause will be something like (13), where the verb forms a sequence with the morphemes heading the finite clausal projection (I(nflectional) Phrase) and the projection specifying clause type (C(omplementizer) Phrase).

(13) Korean Tensed Interrogative

<table>
<tr><td>Mica</td><td>ka</td><td>ecey</td><td>ppalli</td><td>muncey</td><td>ul</td><td>haykyelhay-]$_{VP}$</td><td>-ss-nun-]$_{IP}$</td><td>ya]$_{CP}$</td></tr>
<tr><td>Mica</td><td>NOM</td><td>yesterday</td><td>fast</td><td>problem</td><td>ACC</td><td>solve-</td><td>PAST-ADNOM</td><td>Q</td></tr>
</table>

'Did Mica solve the problem fast yesterday?'

Whether the syntactic units in (13) are composed by transformational operations (Choe 1988; Koopman 2005), or by a post-syntactic morphological operation (Sakai 1998), the crucial properties of a representation like (13) are that (i) a higher level of structure than VP is present, and (ii) elements other than the heads in (13), such as

adverbs, may not break up the sequence.[9] The consequence of this representation is that adverbs following the verb must also follow the highest clausal projection, CP in (13). Such adverb positioning is perfectly possible in "strict" type III languages (14), but it has the interpretation of a Right Dislocation or "afterthought" construction, as predicted by (13), because it involves adjoining adverbs not merely to the verb phrase, but the entire clause.

(14) Mica ka muncey ul phul-ess-nya]$_{CP,}$ ecey / ??ppalli.
Mica NOM problem ACC solve-PAST-Q yesterday fast
'Did Mica solve the problem, yesterday/?fast?'

Summarizing, then, Universal 7 falls out as a derivational generalization in any framework which maps the verb in strict type III languages to a surface position in or directly adjacent to the heads of higher clausal structure.

Universal 10. Question particles or affixes, specified in position by reference to a particular word in the sentence, almost always follow that word. Such particles do not occur in languages with dominant order VSO. (Greenberg 1963: 82)

Universal 10 applies to yes/no (polar) questions. Greenberg (1963: 80–82) observes that question particles may be fixed by position in the sentence as a whole (initial or final), or by reference to some particular constituent, most often the verb, or the emphasized constituent in the question. Universal 10 applies to the latter case, and has two parts. Greenberg finds that the first part holds for thirteen of fourteen languages which have Q particles of this type in his sample. He reports that in one, Yoruba, the particle precedes the constituent in reference to which its position is fixed (1963: 106, note 12), but this appears to be an error.[10] The second part is claimed by Greenberg to be absolute: it states that Q particles of this type do not occur in VSO languages.

In generative treatments, initial and final Q particles have typically been analyzed as complementizers, the highest head in the clause. Some linguists have analyzed the type of Q particle addressed by Universal 10 in a similar fashion: these particles can be analyzed as heading a lower functional projection; the item by which the position of the particle is fixed then moves immediately to its left (see, for example, Julien 2003: 23 for Turkish). There are several subcases of this: in a common one, as noted by Greenberg, the focus of the polar question moves into the specifier of the projection headed by the particle. As observed by Dryer in Haspelmath et al. (2005: 375), in such

[9] In principle an approach of this sort would be available in LFG as well, exactly parallel to Sadler's (1997) and Bresnan's (2001) treatment of VSO languages, where the inflected verb, corresponding to the predicate in f-structure, is inserted into a higher structural head position such as Comp(lementizer) or Infl(ection) in c-structure. In practice, LFG treatments of strict type III languages such as Sells (1995) have tended to adopt the position that such structural categories do not exist in these languages. This makes it unclear why the pattern in (14), for example, behaves like Right Dislocation, rather than right adjunction to VP.

[10] Yoruba has polar Q particles both in final position, as noted by Greenberg, and initial position (Bamgbọṣe 1966). The particle *ni*, which functions as a Q marker in final position, may also appear in second position, but in this position it is a focus marker, not a Q marker (Adéṣọlá, to appear). I am grateful to Victor Manfredi for clarification of the Yoruba facts.

languages there is normally an alternate strategy for marking questions with no focus; often, in this case, the same particle occurs in peripheral position, as in (15–16) from premodern Japanese.

(15) Premodern Japanese polar question particle *ya*

[$_{FOCP}$ [Tatu no kubi no tama] **ya** [$_{VP}$ torite ofasitaru]]
dragon GEN head GEN jewel Q taking came
'Did (he) bring the gem on the dragon's head?' (*Taketori monogatari* c. 900)

(16) [[Ware wo ba sir-azu] **ya**]?
I ACC TOP know-not Q
'Don't you know me?' (*Ise monogatari* c. 900)

The analysis in (15–16) represents *ya* as a phrasal head in both patterns, in construction with the focus or range of the question in both (in (15) 'gem on the dragon's head'; in (16) the entire clause).

This treatment of Q particles makes two predictions with respect to Universal 10 as stated by Greenberg. First, it predicts that when the position of the particle is determined by a particular constituent, the particle should follow. This follows from the principle that specifiers precede their heads, a central consequence of Kayne's (1994) theory of the relationship between phrase structure and linear order, but widely accepted outside that framework as well (cf. section 10.2).

Second, the treatment of polar question markers as phrasal particles concurs with Greenberg that such particles should not occur in VSO languages—except in one specific configuration. This consequence follows from the Head Movement Constraint (HMC; Travis 1984: 131), which disallows movement of one head (such as a verb) past another. Recall from our discussion of Universal 6 above that VSO order is derived by movement of the verb to the highest clausal head position. Movement of the verb over a polar Q particle in a lower head position would violate the HMC, as shown in (17a). The specific configuration allowed by this account (but not found by Greenberg) is a VSO language where the Q particle itself occupies the highest position in the clause, and the verb moves and adjoins to it.

(17) a. *[v. . . . . [$_{FOCP}$ Q . . . [$_{VP}$ . . . t_V . . .]]]
b. [v+Q . . . [$_{VP}$ . . . t_V . . .]]

In a sample of 777 languages, Dryer in Haspelmath et al. (2005) finds 8 languages with polar question particles whose position is neither initial nor final.[11] Of these eight, one, Niuean, is VSO. Niuean exemplifies exactly the configuration in (17b): the polar Q particle follows the clause-initial verbal complex (Seiter 1980: 25–26).

[11] Under the analysis of Q particles as phrasal heads, second-position particles are a subcase of clause-initial particles, where the first constituent is moved into the specifier of the projection headed by the particle or a phonological operation reorders the first word and the particle. Dryer reports 272 instances of final polar Q particles, 118 initial, 45 second position, and 8 "other position" (2005d: 374). The small number of "other position" exemplars reflects the fact that if language allows both "other" and initial or final position, the latter was chosen for coding as "neutral".

(18) Tohitohi k-e kapitiga haau? (Seiter 1980: 26)
write Q-ABS friend your
'Is your friend writing?'

From a derivational standpoint, then, Universal 10 can be revised as follows:

(19) Question particles or affixes, specified in position by reference to a particular word in the sentence, always follow that word. Such particles occur in languages with dominant order VSO only after the initial verb.

Universal 11. Inversion of statement order so that verb precedes subject occurs only in languages where the question word or phrase is normally initial. This same inversion occurs in yes–no questions only if it also occurs in interrogative word questions.

(Greenberg 1963: 83)

Only the first part of Universal 11 is derivational: it entails that subject/verb inversion in wh-questions will not occur when wh-movement does not. Inversion as a question marking strategy is genetically restricted to Indo-European and Uralic (Siemund 2001: 1025);[12] nevertheless Universal 11 appears to be a true substantive derivational universal. Evidence for this can be seen in the subset of languages with wh-movement and inversion in questions which have the partial wh-movement pattern in (20).

(20) German partial wh-movement
a. Was glaubst du [mit wem Maria jetzt spricht]? (McDaniel 1989)
What think you with whom Maria now talks
'With whom do you think that Maria is now talking?'
b. *Glaubst du [mit wem Maria jetzt spricht]?
think you with whom Maria now talks

(20a) is a well-formed example of partial wh-movement, with a dummy wh-word at the left margin of the main clause, inversion, and the semantically contentful wh-phrase moved to the left margin of the embedded clause. We might expect inversion alone to suffice to mark the matrix scope of the question, but as (20b) shows, the result is ill-formed: cross-linguistically, partial wh-movement always appears to require a dummy wh-word on the margin of the main clause, as predicted by Universal 11.

Universal 19. When the general rule is that the descriptive adjective follows, there may be a minority of adjectives which usually precede, but when the general rule is that descriptive adjectives precede, there are no exceptions. (Greenberg 1963: 87)

Universal 20. When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or the exact opposite. (Greenberg 1963: 87)

[12] Ultan (1978: 222) also includes Malay in his inventory of languages with inversion in polar questions, but this is an error: although Malay allows predicate fronting, this is a focus construction independent from interrogatives (Kader 1976). I am grateful to Edith Aldridge for clarification of the Malay facts.

A derivational account of Universal 20 is proposed by Cinque (2005), based on an assumed hierarchical universal: that the underlying order of elements in an extended nominal projection is as in (21):

(21) [Determiner [Number [‘Indirect’ Adjective Phrase [‘Direct’ AP [NP]]]]]

The main derivational constraints in Cinque’s approach are: (i) movement is only to the left; (ii) NP or the phrase that contains it may move, but the other categories (Determiner, Number, AP) do not move independently of NP. NP-final configurations (the second part of Universal 19, the first part of Universal 20) result when nothing moves at all. The two orders referred to in the second part of Universal 20 are derived by fronting the NP alone within the larger nominal projection (preserving the relative order of modifiers), or by a “roll-up” derivation, where NP moves to the left of AP, NP+AP moves to the left of Number, etc., reversing the relative order of modifiers.

Universal 19 also can be stated as a derivational generalization referring to the hierarchy in (21). Cinque (2003) characterizes two kinds of languages which allow adjectival modifiers on both sides of the nominal head: (a) languages like Italian, where (i) NP normally fronts around “Direct” AP modifiers, and (ii) the constituent [“Direct” AP [NP]] always moves around “Indirect” AP modifiers; and (b) languages like English, where (ii) but not (i) occurs.[13] As we saw above, languages with strict noun-final NP structure, such as Chinese or Korean, maintain the underlying structure in (21). This suggests the typology in (22):

(22) a. (i) NP normally moves around Direct AP and
(ii) [Direct AP [NP]] moves around Indirect AP
b. (ii) only
c. No movement

From this standpoint, Universal 19 is a substantive derivational universal which states that (i) and (ii) may occur in combination (a), or (ii) alone (b), or nothing (c), but that (i) does not occur alone.

Universal 25. If the pronominal object follows the verb, so does the nominal object.

(Greenberg 1963: 91)

Plank (2003) lists this universal as absolute, with a question mark. The ten languages cited by Greenberg in support of Universal 25 all involve weak or clitic pronouns, or instances (e.g., Swahili) where the analysis of object pronouns as bound pronouns or concord markers remains controversial.

Among the many analyses of weak and clitic pronouns in structurally oriented theories of syntax, there are some clear points of consensus. One is that weak and clitic pronouns are non-branching, word or X^0-level categories. Cardinaletti and

[13] Thus, for example, English allows the orders [*every INVISIBLE visible star*] and [*every visible star INVISIBLE*], both with the meaning ‘every visible star which happens at the moment to be invisible’; Italian allows only the order [*le stelle visibili (IN)VISIBILI*] (Cinque 2003: 6). In these examples ‘visible’ is the direct adjectival modifier; ‘invisible’ is indirect.

Starke (1999) refer to this as a "structural deficiency"; more generally, it enables them to adjoin to other word-level categories (as in Holmberg's 1986 account of Mainland Scandinavian object shift), or to be realized as the head of a functional projection containing VP (as in Sportiche's 1996 analysis of Romance object clitics). These structural options, all involving positions to the left of the verb, are not available to full NP objects, which are branching, phrase-level categories. NP objects move to the left of the verb only as a consequence of operations which target objects in general.

In this section I have presented derivational accounts of seven of Greenberg's constituent order generalizations. All of these universals have the form of hierarchical or derivational universals (that is, they reference the relative order of two or more categories in the same syntactic structure), and as all of them reference patterns that involve movement or dislocation in widely discussed syntactic treatments, I have classified them as derivational. Five of these (Universals 6, 7, 10, 20, and 25) follow directly from existing syntactic analyses, although to the best of my knowledge none of the authors of these analyses, with the exception of Cinque (2005), makes direct reference to Greenberg's generalizations. Two more (Universals 11 and 19) do not to my knowledge have existing syntactic accounts, but lend themselves naturally to a formulation as substantive universals within such accounts.

10.6 HIERARCHICAL GENERALIZATIONS

We observed in section 10.3 that Greenberg's original constituent order generalizations include only two that can be defined as hierarchical in the sense of (1b): Universal 1, which describes the relative order of subject and object, and Universal 14, which describes the order of conditionals and their antecedents. While 14 appears to be uncontested,[14] Universal 1 is a famously disproven generalization as a statement about surface order. Derbyshire (1977) and Derbyshire and Pullum (1981) cite numerous examples of object-initial languages, and VOS languages are also well attested from Austronesian and the Americas. Examination of the distribution of the six logically possible orders of subject, object, and verb, however, suggests that Universal 1 might be a valid hierarchical universal; that is, a generalization about underlying order.

Let us take as a point of departure an approach akin to Kayne's (1994) which derives surface word order from underlying SVO through four widely attested "basic" operations: subject raising, verb raising, object shift, and VP fronting. These derive the five surface orders in (23).[15]

[14] It is often suggested that the explanation for unmarked antecedent–consequent order in conditionals is to be found in iconicity. If this were correct, we might expect to find languages where among realis (presupposed) conditionals, *after*-type presupposed conditionals always precede, and *before*-type conditionals always follow the consequent. I know of no such language.

[15] Mark Baker (2001) also presents an account of constituent order typology that predicts the five types in (23), while disallowing OSV. To Baker is due the insight that of the two subject-final patterns, VOS should

(23) Derivation of five surface orders

Surface	Underlying	Operation(s)	Moves
a. SVO	SVO	subject raising	subject out of verbal projection
b. VSO	SVO	verb raising	verb to left of clause
c. SOV	SVO	object shift	object to left of verb[16]
d. VOS	SVO	VP fronting	VP to left of clause
e. OVS	SVO → SOV	object shift, VP fronting	

The four operations in (23) indeed are attested in English (23c, object shift, at an earlier period of the language). Except for subject raising, however, the operations are optional in English; in other languages they are basic in the sense that they derive the word orders that typologists have identified as unmarked.

The basic operations in (23) leave only OSV order underived. The antiderivationalist skeptic may object, why not simply add a fifth operation of object supershift, which moves the object to the left of the subject? The answer to this objection is that the two operations displacing NPs in (23) have very specific characteristics: they appear to displace argument NPs to positions where they continue to function as arguments (argument positions in the sense of Chomsky 1981). Displacement of non-subjects to the left of subjects is indeed robustly attested, but it targets non-argument positions: positions associated with questioned, focused, or topicalized constituents. In identifying "unmarked word order", it is precisely argument positions that typologists seek to specify.

On this approach, then, OSV word order can only be derived by a non-"basic" operation, shifting the object directly to the left of the surface position of the subject. Dryer in Haspelmath et al. (2005) identifies four OSV languages: Nadëb (Makú), Tobati (Oceanic), Warao, and Wik Ngathana (Pama-Nyungan). This is a notably smaller number than the other O>S types, VOS (26) and OVS (9). Further inspection of the status of these four languages raises questions about the validity of the OSV analysis.

Nadëb is syntactically ergative (Martins and Martins 1999: 263, citing Weir 1984: 89–91). Basic constituent order is determined by the placement of the absolutive or S/O argument: SV or VS in intransitives, OAV or AVO in transitives.[17] Nadëb constituent

pattern with the two other VO types, while OVS should pattern with SOV (2001: 127–128). However, Baker does not derive basic-word-order variation through movement operations. One consequence of this is that he is unable to explain the relative rarity of OVS languages (nine in Haspelmath et al.'s (2005) sample). In (23), OVS results from the co-occurrence of two relatively rare basic operations, object shift and VP fronting.

[16] As noted in section 10.2, strictly speaking, object shift derives SOVX order. Greenberg's "strict subtype" of SOV languages, where all arguments and adjuncts precede the verb, requires something like the kind of derivation in (5) (Kayne 1994). But object shift appears to be the "basic" operation that interacts with VP fronting to derive OVS languages; thus Hixkaryana is OVX, not OXV.

[17] I use here the terminology introduced by Dixon (1979), to make the point that the relative order of major constituents in Nadëb should not be characterized in terms of subject, object, verb. A denotes transitive subject, S intransitive subject, O object.

order is thus best described as Abs-X-V, alternating with X-V-Abs.[18] Wik Ngathana is also ergative (Sutton 1978: 285–286).

In the two remaining languages, the OSV accounts contradict earlier SOV analyses. Donohue (2002: 198) suggests that a recent change SOV>OSV may have taken place in Tobati, since Cowan (1952) reported the language as SOV. But Donohue's own data is also predominantly SOV in transitives with both S and bare O present (seventeen of twenty-seven examples). The examples of clause chaining in (24) suggest that OSV order in Tobati may not reflect the basic argument position of O.

(24) Tobati (Donohue 2002: 191)
a. Man har-ad rom-ra yar.
bird person-ALL see-SEQ fly
'The bird saw the man and then flew off.'
b. Man har rom-ra yar.
bird person see-SEQ fly
'The man saw the bird and then it flew off.'

What is interesting about (24) is that clause chaining is controlled by the first NP in the clause, regardless of whether it is subject (24a) or object (24b). Donohue suggests (p.c.) that allative marking on the object with -*(a)d* as in (24a) may in fact be the normal accusative pattern; bare objects (particularly definites and generics, judging from Donohue's data) are pervasively topicalized, and clause chaining as in (24) is controlled by the clause-initial topic. If this analysis is correct, the basic argument position of S and O in Tobati is SOV.

Similar considerations come into play for Warao. Romero-Figueroa disputes the earlier claim of Osborn (1962) that Warao is SOV. Like Donohue, Romero-Figueroa relies on consultants' translations of examples with reversible subject and object, specifically:

(25) Yatu hua mi-ya. (Osborn 1962: 260; Romero-Figueroa 1985: 127)
you Juan see-PRES
Osborn: 'You see Juan.'
Romero-Figueroa: 'Juan sees you.'

[18] Ergativity is also relevant to the classification of the nine OVS languages described in Haspelmath et al. Six are South American: Hixkaryana and Tirio (Cariban), Selknam (Chon), Cubeo (Tucanoan), Asuriní (Tupian), and Urarina (Urarinian). This suggests that the two basic operations associated with OVS order above, object shift and VP fronting, co-occurred (viewing them now as diachronic innovations) in a fairly circumscribed area on the eastern half of South America. From this standpoint, the three Eastern Hemisphere OVS languages are outliers. But they are outliers syntactically as well: all three have SV order in intransitive clauses, while all six South American languages are VS; the sole African language (Päri, Nilo-Saharan) is ergative, while of the two Australian languages, one has split person marking (Mangarrayi) and the other (Ungarinjin, Wororan) is described as accusative. SV order indicates that all three have a derivation quite different from the OVS South American languages, and at least the first two may be better described as having Absolutive-VX basic order.

However, difficulties arise with this kind of test in languages which typically topicalize discourse-presupposed NPs (such as discourse participants). In fact, one of Romero-Figueroa's primary sources, Vaquero (1965), directly contradicts Romero-Figueroa's claim about the unmarked order of reversible S and O. Vaquero states that when the meaning of the sentence would otherwise be unclear, the order is SOV (1965: 144). A survey of transitive clauses with both unmarked S and O present in Vaquero's narrative texts reveals an overwhelming preference for SOV: forty-six examples of SOV versus four of OSV.[19] Virtually none of the examples of initial S is discourse-new or otherwise highlighted, casting doubt on Romero-Figueroa's claim that SOV is derived by focus fronting of S. Instead, as with Tobati, OSV appears to be the product of pervasive topicalization of definite and generic objects.

I have argued in this section that Greenberg's Universal 1 is a valid hierarchical universal, that is, accurate as a generalization about underlying constituent order. Crucial to this argument has been a claim that there is a high degree of congruence between what I have called "basic" operations—operations that permute subject, object, and verb without affecting the argument status of the former two constituents—and the informal criteria that typologists use to identify basic word order.

10.7 CONCLUSION

Since Greenberg (1963), work on constituent order typology has tended to privilege the status of cross-categorial generalizations, in both "formal" (particularly transformational) and "functionalist" frameworks. In the former case, the longstanding appeal of X'-theoretic generalizations has surely played a role. If the argument developed in this paper is correct, cross-categorial generalizations of the X'-theoretic type are not the product of an imperative in Universal Grammar to favor, for example, consistency in head-complement order across categories, nor are they the product of pressures from language processing. Instead they are the result of well-documented patterns of

[19] I am indebted to Anne Gagliardi for the analysis of Vaquero's texts. Vaquero's exact statement is the following (1965: 144):

a. *Regla general:* El sujeto precede inmediatamente al predicado; v.gr.: Oko naruya diana: Nosotros nos vamos ya. Domu koitaya: El pajarito canta. Kuaimo a araotuma nanakanae: Los habitantes de arriba bajaron.

b. *Excepciones:*
Siempre que el sentido de la oración pudiera resultar equívoco o dudoso, el sujeto se antepone al complemento directo; v.gr.: Tai yatu yewereae: El os castigó. Invirtiendo el orden resultaría todo lo contrario: Yatu tai yewereae: Vosotros le castigeis. Tobe yatu nae: El tigre os mató. Yatu tobe nae: Vosotros matasteis al tigre.

The examples that Vaquero provides in (a) are all intransitive, while those in (b) are transitive, and make clear that with reversible subject and object, the subject precedes.

language change, which are common, but not necessary or even favored, from the standpoint of acquisition or performance.

In an insightful discussion of the place of universals in syntactic theory, Newmeyer (2005) points out that cross-categorial generalizations such as the Head Parameter are a continuation of a longstanding effort to incorporate markedness considerations, in this case about syntax, into formal grammar. The argument in this paper that cross-categorial markedness generalizations such as the Head Parameter are not statements about synchronic grammar, but rather products of language change, can be compared to the critical assessments of markedness principles in synchronic phonology and morphology in the papers by Blevin and Garrett in this volume.

At the same time, this paper has left a very robust role for syntactic universals. I have argued that syntactic universals are to be situated in the exact areas of central interest to generative and indeed structuralist syntacticians: the underlying hierarchical arrangement of constituents and the constraints on mappings between levels of representation. These last were certainly not Joseph Greenberg's main areas of concern in his research on constituent order generalizations, but I have argued that his most robust universals reflect them.

11

Emergent Serialization in English: Pragmatics and Typology

Paul J. Hopper
Carnegie Mellon University

11.1 INTRODUCTION

We can say of change in language that it is constant, gradual, and always driven by discourse. The opposite of these postulates would be that change is stadial, discrete, and driven by sentence- and clause-level formal conditions. While the latter set of postulates underlies, implicitly or explicitly, much work in diachronic linguistics, there is a wealth of research that points to the need to study change in its broader discourse context, that is, through usage. When in widely separated languages change results in closely comparable types of constructions, we should look closely at the common discourse conditions that might be conducive to the emergence of such constructions.

This functionalist-diachronic approach to constructions is hardly novel. Yet the study of language typology is to an overwhelming extent pursued at the level of the single, isolated sentence presented in sets of translation equivalents, without regard to such aspects as frequency of usage or analysis of contextual meaning. It posits fixed states characterized by rigidity and stability, and a taxonomic attitude that assumes languages can ultimately be matched without remainder to categories such as *pro-drop language*, *classifier language*, *SVO language*, *serializing language*, and so on. Quite often, sentence-level grammatical arguments are adduced in order to eliminate troublesome alternants, the appearance of uniformity being sustained in the face of manifest intra-language variation by means of descriptors such as “dominant” (e.g., word order), “predominant”, “underlying”, “basic”, and the like. One undesirable result of this limitation is that through it we are denied a motive for studying discourse patterns that might reveal the germ of a typologically relevant construction in the usage of speakers of languages not normally known for that construction. Yet often, types of construction that are central and robust in some languages may be identified in a weaker and more rudimentary form in others. As Bybee notes (this volume),

"synchronic universals ... whether they be substantive or formal, can be thought of as emergent: they are not given apriori, but arise through the interaction of other more basic processes". This observation suggests that we should ask not *whether* a language possesses this or that typological characteristic, but *to what degree*, and points in turn to the methodology of actively searching for the weak counterparts of robust constructions in a range of languages and especially of examining carefully their discourse distribution. It is only through such a study that we can hope to explain why some types of construction are so persistent in widely disparate languages and to understand the historical conditions under which such constructions might arise.

The typological feature in focus in this paper is verb serialization, the use of two or more verbs in sequence inside the same clause and forming part of the same predication. More narrowly, a use of the verb *take* in English discourse is described which strikingly recalls that of its translation equivalent in verb-serializing languages. While marginal serializing phenomena have been noted for English (e.g., in Pullum 1990; Durie 1997; Hopper 2002), the studies concerned have usually been restricted to intransitive combinations such as *go get*, or coordinated pairs like *try and find*. The possibility of a transitive serializing construction, where the two verbs are not contiguous but are separated by a direct object, in English seems less compelling. While there has been much discussion of questions of reanalysis within the general pattern of already serialized verbs (emergent case marking, verbal prepositions, verb compounding, auxiliation, etc.), the question of what earlier stages might have looked like has all but been ignored. Yet patterns of verb serialization involving *take* are widespread in the world's languages. Once we have set aside appeals to magic in the form of "parameter setting", it is clear that the answer to this and other questions about typology must lie in a combination of common discourse needs and historical change. The possibility of discourse factors that might have led to the semantic bleaching of *take* and its subsequent grammaticalization as a case marker that is so frequent in verb-serializing languages calls for a closer investigation.

11.2 TYPOLOGICAL FEATURES OF SERIALIZATION

The serial verb construction was first identified in African languages. Although serial verbs have since been discovered in many parts of the world, it was their prominence in certain West African languages that first compelled the attention of linguistic theorists. For those whose conception of a sentence or clause demanded one and only one verb, serial verbs were mysterious, and presented a challenge to formulate analyses that would fit them into the single-verb clausal template demanded by syntactic theory. A wealth of research has led to an appreciation of the complexity of verb serialization. Indeed, some linguists, especially those approaching language from a

generative perspective, have disputed the existence of the type altogether. But a broad consensus has arrived at something like the following list of defining properties of the prototypical verb serialization construction (Sebba 1994; Durie 1997; Pawley and Lane 1998; Aikhenvald 1999):

1. Each verb is in principle tensable (i.e., none of the verbs is non-finite).
2. The verbs have the same values for tense, aspect, and modality.
3. There are constraints on inflectional possibilities in the verbs.
4. No overt characteristics of clause boundaries (such as a conjunction, intonation) are present.
5. While there are two verbs, only one event is predicated. (Serialization is *monopredicative.*)
6. There are constraints on core arguments: either the two subjects are identical, or the undergoer of the first verb becomes the actor of the next.
7. Negation and adverbs whose scope is the clause have scope over both verbs.

Before moving to consider the verb *take* as a possible English candidate for serialization, we should add two comments that will later become relevant. To 4., the constraint that there may be no overt signals of clause boundaries, we might ask if this constraint as stated is not too strong. In a number of European languages, constructions closely akin to serialized verbs exist in which two clause-like sequences expressing a single event are obligatorily linked by the weak coordinator translated with English *and*. Kuteva, in her monograph on auxiliation (Kuteva 2001), cites instances from Bulgarian and Danish, among others, where verbs such as *sit*, *stand* function as quasi-auxiliaries, as in (1)–(3).

(1) Bulgarian (Kuteva 2001: 70)
a. Drexite sedjat i săbirat prax
clothes.DEF sit and gather dust
'The clothes are gathering dust all the time'
b. Krepostta stoi i se ruši s vsjaka izminata godina
fort.DEF stand and REFL go.to.ruin with each past year
'The fortress is falling to ruin from year to year'
c. Trionăt leži i răždjasva v mazeto
saw.DEF lie and get.rusty in cellar
'The saw is getting rusty [lies rusting] in the cellar'

(2) Middle Dutch (Kuteva 2001: 69)
De steden staen dagelicx ende vervallen
the cities stand daily and go.to.ruin
'The cities are falling to ruin from day to day'

(3) Danish (Kuteva 2001: 47, citing Braunmüller 1991: 103)
Han sidder og spekulerer over fremtiden
he sits and speculates over future.the
'He continually speculates over the future'

Here, Danish *sidder* has no semantic trace of 'to sit', but is purely aspectual. Other verbs can be used in this way, such as *gå* 'go', *stå* 'stand', and *ligge* 'lie'. While these verbs have been considered as emergent aspectual auxiliaries, they obey all the constraints mentioned above for serial verbs except for that concerning the presence of an overt signal of clause boundary (Bulgarian *i*, Middle Dutch *ende*, Danish *og*). In examples (1c) and (2), it is instructive to note the position of the adverbials (Bulgarian *v mazeto* 'in the cellar', Middle Dutch *dagelicx* 'daily'), reflecting the different degrees of grammaticalization of the double verb construction. In Bulgarian the two verbs appear as a unit, with the adverbial that means 'in the cellar', which belongs semantically to *leži* 'lie' rather than with *răždjasva* 'rust', placed outside the two conjoined verbs. In Middle Dutch, on the other hand, the adverb *dagelicx* 'daily' appears after the first verb and separates the two verbs. According to Kuteva's source, this construction did not survive into modern Dutch, and so example (2) represents a stage prior to full grammaticalization.

It is surely significant that such double verbs obligatorily linked by *and* are a typically European feature,[1] for Europe is one of the few major linguistic areas in which verb serialization is not found. The availability of a neutral, omnipurpose coordinator comparable to *and* would seem at first sight to inhibit the emergence of serialization as defined by the full set of features 1–7 above, since *and* can always be seen as a marker of a clause boundary. However, many languages do not conjoin clauses with a special obligatory conjunction like *and*, but use participles, a preposition meaning approximately 'with', or simple parataxis, or, sometimes, have recently introduced new clause coordinators through Western influences; often these are borrowed forms (on the typology of coordination, see Stassen 2000; Haspelmath 2004c). The absence of an *and* word in true serialization could therefore simply be a manifestation of the lack of a general coordinator in these languages, and we may be justified in overlooking the presence of *and* when seeking evidence of serialization.[2] A recent paper on the Southeast Asian language Karen (Lord and Craig 2004) shows how in that language serialized verbs lacking a coordinator are viewed as single events, and contrast with overtly coordinated ("concatenated") sequences of verbs that are viewed as distinct events, as in (4).

(4) Sgaw Karen (Lord and Craig 2004: 365)
  a. 'əwɛ hɔ dɔ' lɔxi
     3SG cry and fall-down
     'He cried and fell down'
  b. 'əwɛ hɔ lɔxi
     3SG cry fall-down
     'He collapsed from grief'

[1] Though by no means exclusively so. Stassen (2000) is now the *locus classicus* for the general typology of coordination.

[2] Aikhenvald asserts that, for languages that mark subordination or coordination explicitly, "[s]erial verbs are monoclausal and allow no markers of syntactic dependency between their components. This distinguishes them from subordinate or coordinate clauses" (Aikhenvald 1999: 470). But she also notes (ibid.) that exceptions can be found to all the putative defining properties of serial verb constructions.

Here the coordinator is syntactically optional and semantically significant. Because in English the coordinator is not optional, a claim for verb serialization would at least sometimes have to be justified on semantic or pragmatic rather than structural grounds; it would have to be shown that in some contexts it was necessary to interpret a concatenated sequence of verbs as a single event even though in other contexts it represents a sequence of distinct events.

A second comment has to do with item 3. in the list of properties, constraints on inflection in the two verbs. In serializing languages where tenses, moods, and aspects are morphologically marked, such morphology is typically restricted in serialized verb sequences in such a way that all verbs are uniformly marked. The number of paradigmatic possibilities in serialized verbs may also be more restricted than in single verbs. Typically, too, in serializing languages noun phrases do not carry morphology indicating their argument status.[3] The end result of these restrictions is that serial verb constructions are characterized by a certain amount of morphological simplification. From this point of view, the examples cited in (1), (2), and (3) above from Bulgarian, Middle Dutch, and Danish respectively should be seen in the context of languages that have lost earlier morphological complexity, to which list English should be added. However, if serialization is in an incipient stage, we might expect any of the features of serialization in 1–7 to be manifested, not absolutely, but relatively. This might mean, for example, that there would be a *tendency* for the verb forms that are used in serialized two-verb constructions to be less complex morphologically than those same verbs used in other contexts. Thus the English *turn around and* construction (Hopper 2002) occurs in both serialized and non-serialized forms, as in (5)–(6).

(5) Unless you have well- wellies on . . . you turn round and come back

(6) you ask ʻem to lend you a fiver and they might <laughs> turn round and tell you to sod off

In (5) the speaker is referring to a distinct physical event of turning, evidently because progress along the path has become impossible. Turning and coming back are presented as distinct events potentially having different scopes (it would be possible to turn round but not come back, one could turn round quickly but come back slowly, one subject could turn round and another come back, etc.). But in (6) no action of turning is suggested, and *turn round and tell* are presented as a single event—a conceptual unit, as Lord and Craig (2004) put it—with some special modality (see Hopper 2002 for discussion of this particular construction). In this case there is a slight but unmistakable favoring of the forms of *turn* that lack an overt inflection. Table 11.1 (from Hopper 2002) shows that by comparison with the overall distribution of the lemma *TURN*, in which fewer than half of the forms consist of *turn*, in the

[3] According to Aikhenvald (1999: 474), "[s]erializing languages have little or no case-marking . . . [m]ost serializing languages are typologically head marking or else neither head nor dependent marking". However, Aikhenvald finds in her Amazonian data counter-examples to these and a number of other generalizations that have been made about serial verb languages.

TABLE 11.1. Forms of the lemma *TURN* in *turn around and* (Hopper 2002)

Percentage of form of *TURN* in *turn around and*			Percentage of all forms of *TURN*		
turn around and	66	(n = 130)	*turn*	46	(n = 1002)
turned around and	22	(n = 44)	*turned*	32	(n = 705)
turning around and	9	(n = 19)	*turning*	12	(n = 271)
turns around and	3	(n = 5)	*turns*	10	(n = 219)
Totals	100	(n = 198)		100	(n = 2197)

serial construction the base form (or general present)[4] *turn* occurs two-thirds of the time.

The end result of such a process may be discerned in the comparable modal construction with *try*, where *try and* has already reached the point where all forms of the lemma *TRY* are excluded except the base form *try* (Quirk, Greenbaum, et al. 1985: 979; Pullum 1990; Biber et al. 1999: 738).

11.3 THE VERB *TAKE* IN SERIALIZATION

A frequently encountered phenomenon in serial verb constructions is the appearance of some such construction as the following from fifteenth century (Middle) Chinese:

(7) Middle Chinese (Sun 1996: 91; tones unmarked in source)
jiang dian-zhu de-si lai
take store-owner hit-dead CR (=current relevance)
'[Someone] has beaten the store owner to death'

A verb of general meaning conventionally translated as 'take' acts both to transitivize and to redistribute the NP arguments. The modern equivalents of such sentences have *bà*, whose earlier meaning was 'take', in place of *jiang*. Increased transitivity is a typical semantic accompaniment of the use of *bà*. Sun notes, for example, that a *bà*-marked object is incompatible with an assertion of non-completion (1996: 55):

(8) Chinese
a. ta he le tang le, keshi mei he-wan
3 drink ASP soup ASP but NEG drink-finish
'He has eaten the soup but did not finish it'
b. *ta ba tang he le, keshi mei he-wan
3 BA soup drink ASP but NEG drink-finish
'He has eaten the soup but did not finish it'

Contrasts of this kind point to a transitivizing function for *bà* (on this point see Hopper and Thompson 1980: 274–275). Analogous facts can be adduced for other

[4] The two are conflated in the present paper, because the overt absence of a suffix is precisely what is at issue. Strictly speaking, though, the base form would not include finite non-suffixed forms as in *they turn*. For discussion of this point, see Pullum (1990).

serial verb languages. Carol Lord in her book on verb serialization in West African languages points to numerous constructions involving a verb translated as *take* in Twi, Idoma, Nupe, Dagbani, Gwari, Engenni, Awutu, and Vagala, as well as the New Guinea language Kalam and in Choktaw (Lord 1993: 65–138). Examples of some of these are presented later in this paper.

While we must be cautious in accepting the automatic translation of all these verbs with a single English verb *take*, there are some interesting parallels between the distribution of the verb *to take* in contemporary English discourse and the serial verb translated as *take* in these languages.

11.4 THE ENGLISH HENDIADIC *TAKE NP AND* CONSTRUCTION

Grammatical uses of *to take* in English have not received much attention as a class, and yet they are central in a number of constructions that substitute for or supplement lexical verbs, including well-known (and less neglected) expressions like *take a shower*, *take offense*, *take a sip*. What might be called a syntactic use of *take* occurs in examples of the following kind. The example in (9a) is hypothetical.

(9¢)[5] a. They enlarged the same design as before by including a library and a gymnasium.
b. They *took* the same design as before *and enlarged it* by including a library and a gymnasium.

A search of several English corpora revealed many examples of this construction in actual discourse, so that fictional examples are not necessary.[6] The example in (10) is typical.

(10) Then, what I would also do is take the number A under that list that really expands this and move it down to the last on the list instead of the first. (CSPAE)

For brevity's sake this construction will be referred to as the *take NP and* construction. It is hendiadic in the sense that a single meaning is distributed over two constituents (Hopper 2002). The term *hendiadic* (Greek *hen dia duois* 'one through two') refers to a figure in which a pair of forms joined by *and* is understood to

[5] The warning symbol ¢ (for "constructed") has been used to mark an example as having no textual warrant.

[6] The following corpora were used in whole or in part as data sources in this paper:

Cobuild: *Collins-Birmingham University International Linguistic Database*. Unidentified citations are from this corpus.

CSPAE: *Corpus of Spoken Professional American English,* by Michael Barlow, available from Athelstan <http://www.athel.com/cspa.html>

SB-CSAE: *Corpus of Spoken American English*. John Du Bois et al., Department of Linguistics, University of California at Santa Barbara. (The Santa Barbara Corpus)

SWITCHBOARD: Telephone Data Corpus, Linguistic Data Consortium, University of Pennsylvania

have a meaning paraphrasable by a single constituent, such as *rock and roll*, *nice and warm*. It contrasts with *synthetonic*, referring to two distinct forms with no merger of meaning, such as *black and white*, *warm and dry*. The *take NP and* construction involves two grammatical clauses, the first of which contains a transitive verb *take* and a direct object. The second clause verb is usually also transitive, and its object is typically a pronoun referring to the object of *take* in the first clause. There is thus an interesting symmetry between the coordinated clauses, in that *take* in the first clause refers forward cataphorically to the lexical transitive verb in the second clause, without which *take* is meaningless, and *it/them* in the second clause refers back anaphorically to a noun phrase in the first clause. The effect is that of a transitive verb construction such that *the lexical verb and its object are in different clauses*, that is, the verb and the object are grammatically distributed over two syntactic clauses, with the object first and the verb second. Adverbial complements after the direct object are also very characteristic of the construction, as in the example given. Very often this is a prepositional phrase. The hendiadic *take NP and* construction is well represented in the corpora; so richly, in fact, that it is surprising that it has not previously been recognized as a grammatical construction. For example, the *Oxford English Dictionary*'s (2nd edition) massive entry under *take (v.)* gives no examples of this hendiadic use. No mention of the construction is made in either of the standard reference grammars of English, Quirk, Greenbaum, et al. (1985) and Biber, Johansson, et al. (1999). As examples to be presented will show, *take NP and* prefers monologic contexts where extended arguments are being made; yet it is characteristically a spoken rather than a written language construction.

11.4.1 Synthetonic versus hendiadic *take NP and*

No doubt one of the reasons why the *take NP and* construction has not been recognized is that synthetonic and hendiadic versions are often homophonous. In the examples in (11)–(12), the verb *take* unmistakably refers to a distinct event:

(11) She felt something—someone—take her hand and squeeze it gently

(12) Oh take one and pass it on

In (12), it would be possible to "take one" but not "pass it on". The examples in (13)–(14), however, present a single event distributed over two verbs:

(13) And unfortunately we're going to have to take all these people and squish them into a church that seats four hundred with an overflow another four or five hundred

(14) You don't think this is a bit of a generalization here David do you just you're just taking the odd one and making it into the norm

In such utterances, the verb *take* and the next verb are not independently negatable, and could not be modified by different adverbs. When necessary, the adjective *hendiadic* will be used to distinguish examples like (13) and (14) from examples like (11) and (12). We are justified in considering the hendiadic sequence of *take* and a second

verb as a *construction* in the sense in which this term has recently come to be used by such linguists as Fillmore (1988), Langacker (1987), and Goldberg (1995): a closed schematic formula in which limited lexical substitutions can be made, and which is grammatically dispensable, that is, the language could get on just as well without it. I will later question whether it does in fact qualify as a construction in any standard sense.

11.4.2 The elements of *take NP and*

The elements of the schematic formula that defines the *take NP and* construction are as follows, presented in left-to-right order. (There is no evidence that the two clauses are linked by any special suprasegmental feature such as intonation.[7])

1. A subject NP that is common to both verbs;
2. the verb *take*, in any of its forms;
3. a direct object NP that is the target of both verbs;
4. the word *and*;
5. a second verb, which is transitive;
6. an anaphoric pronoun coreferential with (3);
7. an adverbial complement, usually a prepositional phrase.

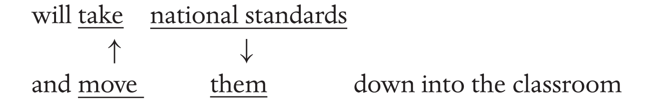
Schematically, we can represent the construction as follows. (The symbol "^" represents concatenation.)

$$\text{NP\^{}}\textit{take}\text{\^{}NP}'_{\text{lex}}\text{\^{}}\textit{and}\text{\^{}V}_{\text{trans}}\text{\^{}Pronoun}'\text{\^{}Adverbial Complement}$$

This formula will be referred to as the *canonical schema* for *take NP and*. Example (15) is a typical representative of the construction in a corpus:

(15) This test . . . will take national standards and move them down into the classroom. (CSPAE)

It can be seen that there is an interesting symmetry in the construction. The direct object NP is resumed anaphorically by a pronoun in the second clause, and the transitive verb of the second clause refers cataphorically to *take* in the first clause, as schematized below.



[7] The four examples of *take NP and* in the audio part of the Santa Barbara Corpus show no special intonation patterns of the two clauses. In two of the examples they are run together under a single contour; in the two others the intonation suggests two quite distinct clauses with an audible pause between them.

Take thus functions as a pro-verb that anticipates, or "projects", the transitivity of the second clause.

There is syntactic evidence that the sequence of verbs in such utterances is understood as a unit.

(16) Other times I'd maybe take half the class and assess them at one task and the other half on a different task later on

In (16) "gapping" lumps together *take* and *assess* for the purpose of ellipsis; that is, the ellipsis makes reference not to *take*, nor to *assess*, but to *take and assess* jointly. Restoring the deletion thus yields (17).

(17¢) Other times I'd maybe take half the class and assess them at one task, and [take] the other half [and assess them] on a different task later on.

The verb *take* is bleached of information content to the point where it can be ignored. The pattern of gapping is exactly the same as with an auxiliary+verb combination, as in (18):

(18¢) Robin can drive the Ford and Leslie the Opel

This equivalence is a further indication that the schema *take NP and* is treated conceptually as a single element. In the example in (19), from written English, the *take* clause and the next conjoined clause are seen to constitute a single event in a sequence of discrete events:

(19) ...bloody handkerchief. He had to get rid of it before the District Attorney showed up in the morning; it had been in the pocket of his pants until now. He got out of bed, pulled on his bathrobe, *took the wadded handkerchief and put it into the pocket of the robe* and softly opened the door to the hallway. The apartment was quiet and dark.

Here the italicized clause, as shown by the punctuation and the use of *and*, constitutes the single item E_3 in the narrative event sequence E_1-E_2- E_3-E_4, where E_2 is *pulled on his bathrobe* and E_4 is *softly opened the door to the hallway.*

11.4.3 Pragmatic and rhetorical functions of *take NP and*

The sequence of verbs not only forms a conceptual unit, it also suggests unified pragmatic and rhetorical functions. One of these is to control the flow of information in the discourse. Without the use of *take NP and*, there would be a surfeit of lexical arguments in the construction as a whole. In (15) we have (not counting the subject, which is usually non-lexical): *national standards*, *move*, *down*, and *classroom*. Without the *take* construction, the sentence would compress these four lexical arguments into a single clause, as in (20).

(20¢) This test will move national standards down into the classroom

The *take NP and* construction permits such multiple lexical arguments to be distributed over two surface clauses, the direct object appearing in the first clause and the

verb and adverbial complement in the second. It can thus be seen as a syntactic strategy in the direction of conformity with the "one clause at a time" principle of information structure, the tendency for parsimony in the amount of significant information that is delivered in each clause (see Pawley and Syder 2000 for a recent statement of this principle).

But this cognitive motivation for *take NP and* is at the most only part of the story. The discourse circumstances conducive to the appearance of multiple new lexical items in a short text are rather unusual. *Take NP and* is rare in casual, face-to-face conversation. It favors contexts, spoken and to a lesser extent written, in which more extensive arguments are being made and thoughts are being marshaled. In (21), an attorney is arguing that public money may not be used to supply a sign language interpreter for one Jim Zobrest, a deaf child attending a religious school:

(21) Lawyer Richardson argued that a sign language interpreter is just as involved in conveying a religious message. 'When you take a public school employee,' he said, 'and ask him to go to work and tell Jim Zobrest that Jesus Christ was the son of God, that is active involvement in the delivery of a religious message.'

The argument is transparently distributed over several clauses, even though the grammatical means are available for bundling them together in a single clause, as in (22):

(22¢) Asking a public school employee to tell Jim Zobrest that Jesus Christ was the son of God is active involvement in the delivery of a religious message.

In (21), *ask NP* has become *take NP and ask him*. But other rhetorical devices are also at work in the task of persuasion. A second-person pronoun has been presented as the subject of *take*, introducing direct and inescapable addressivity into the discourse. *Tell* becomes *go to work and tell*. The subject of the sentence in (21) appears as the finite topic clause *when you take a public school employee and ask him to go to work and tell Jim Zobrest that Jesus Christ was the son of God*. Moreover, the arguments are presented in a protected form that presupposes their factuality, namely the subordinate clause complex introduced by *when*.

Such elaboration of simple messages carries its own oratorical force. To the unaware listener, the enormity of the proposal seems increased with every new verb. A message that could be compressed into a single sentence is spread out over several clauses. Thus it is not uncommon for *take NP and* to be used in combination with other "delaying" constructions like the *go to work and* of (21), or the pseudocleft illustrated in (23)–(24):

(23) What was done here was that if if you take a glass slide and coat it with a a thin layer of a substance which has this property of scintillating when radiation strikes it er in this case zinc sulphide and that's one of the ones

(24) This test, part of what this test will do, I think, will take national standards and move them down into the classroom. (CSPAE)

The elaboration of a simple idea holds the listener's attention a little longer and in so doing may make that idea seem more persuasive. Moreover, the cognitive cost of

a routinized construction like *take NP and* or pseudocleft is minimal. The speaker's mind is freed up and can be devoted to composing the upcoming utterance in the most effective manner.[8]

However, the rhetorical motives for stringing out the delivery of a message include not just the estimated effect on the listener, but the speaker's own social posture in the speaking situation. As long as her utterance can be convincingly prolonged, the speaker has the floor. It is to her advantage to hold off her main point as long as she can do this without losing the attention of her audience and without yielding the floor. The speaker of (10) has at least two objectives. One is to identify a certain place in the report and to recommend moving its number down to the bottom of a list. But she also wants to hold the attention of the listeners in a complex sequence of ideas, and even to focus attention on her words and attribute importance to them by spreading them over two prosodic periods. Putting the verb *move down* and the direct object *the number A under that list that really expands this* in different clauses accomplishes all these goals, since the transitive verb is now in focus. The speaker could have said "move the number A down to the last on the list instead of the first". Instead she introduces the direct object as a complement of the dummy verb *take* and postpones the lexical verb to the next clause: "take the number A . . . and move it down to the last on the list instead of the first". The listener who wants the entire set of ideas, which crucially involves the verb, must wait until the sequence is done, and cannot anticipate the ending and seize the floor with an interruption (see Ochs, Schegloff, and Thompson 1996).

Although found in a variety of genres, including casual conversation, the *take NP and* construction is perhaps especially frequent in institutional contexts such as radio broadcasts, committee meetings, public addresses, and teaching where speakers are conventionally allowed to hold the floor longer than in conversation, and where a rhetorical premium is placed on longer and more complex utterances. It is less common in written English, presumably because the motivations for using it are typically extempore ones involving the speaker's direct social and cognitive assessments of current situations.

11.5 *TAKE NP AND* AS AN EMERGENT CONSTRUCTION

In general, it may seem as if *take NP and* has "arrived" as a fully present and robust English construction with the structural properties and pragmatic functions described

[8] Albert Lord, writing on the technique of oral poets in the Balkans, describes in such terms the tactic of inserting a fixed, routinized expression (formula) into a line under pressure of time while the next line is being mentally organized: "The singer's problem is to construct one line after another very rapidly. The need for the 'next' line is upon him even before he utters the final syllable of a line. There is urgency" (Lord 1960: 54). The formulas "are useful to the singer, for they emerge like trained reflexes" (Lord 1960: 58). For formulaic language in ordinary English discourse, see now Wray (2004).

above. There are more than enough examples of its occurrence in a wide enough array of contexts to establish it as part of the grammar of English to approximately the same degree as more familiar constructions like Passive and Cleft. Yet a close examination of *take NP and* suggests that this impression of stability and robustness may not be the whole story. The schema presented above in defining the construction is here repeated:

$$\text{NP}\hat{\ }\textit{take}\hat{\ }\text{NP}'_{\text{lex}}\hat{\ }\textit{and}\hat{\ }\text{V}_{\text{trans}}\hat{\ }\text{Pronoun}'\hat{\ }\text{Adverbial Complement}$$

This schema is by no means always exactly matched. We will consider examples that "fail" in one or two respects to conform to the canonical schema (for further discussion see Hopper to appear). Sometimes the divergence from the canonical schema is minor, and could be accommodated as a non-significant variant. But there are also cases that stretch the relationship to the breaking point. In what follows the abbreviations C1 and C2 stand for the *take* clause and the follow-up clause respectively.

11.5.1 Pronoun in C2 is not direct object

In the first set of examples that strain the canonical schema, the pronoun that resumes a mention of the direct object of *take* in C1 is not itself the direct object of the C2 verb. There are several possibilities here.

(i) The anaphor in C2 is a prepositional complement

Most commonly the non-canonical construction has the resumptive pronoun as a prepositional complement instead of a direct object, as in (25)–(27).

(25) It was finally decided that he would take your present article and comment on it.

(26) Some people take their current review procedures and build an extra step onto them, that uses what's going on, and that triggers a then more extended review where it seems appropriate. (CSPAE)

(27) Santa Barbara Corpus (Conv. 5)

PAMELA: [But there's] there's me=,
insi=de.
.. That's ... invisible.
DARRYL: ... It's not,
... it's it's n-,
it's it's,
(H) I mean,
(H) what if,
what if you took the same ... spacesuit?
... And you put another spirit into it.
... It would be [a different person

In some instances, for example (25), this type deviates still further from the canonical schema in that the verb of C2 is not transitive.

(ii) The anaphor in C2 is in a subordinate clause

The resumptive pronoun in C2, instead of being the direct object of the C2 verb, appears in a subordinate clause in examples like (28)–(29):

(28) Now the—the premise of this contest was to take the $500 you were given with which to compete and prove that you could spend it more effectively

(29) Build a versatile core wardrobe around your shape. Here we take four GH [*Good Housekeeping*] readers and suggest the clothes that suit them best

(iii) The anaphor in C2 is a possessive pronoun

Here the anaphor appears as a possessive pronoun attached to a lexical noun, as in (30)–(31). The possessed lexical noun of C2 may have one of a number of roles in C2.

(30) Another method is to take a prefix and consult a dictionary for its cognates.

(31) The station took familiar songs and changed their words.

(iv) The C1 direct object appears in C2 as an implied instrumental

In this type, the direct object of C1 is not mentioned at all in C2, but is implied as the tacit instrumental of the C2 verb, as in (32)–(35):

(32) then take a pair of scissors and cut it

(33) then we can take a nice sharp spade and chop them off

(34) The process is a bit like taking a hammer and smashing a clock to find out what makes it tick.

(35) We take Froot Loops and we coax it out of the starting gate.

Since there are parallels for this variant in the serializing languages, the implied instrumental is not as distant from the typological prototype as it might seem. Compare the West African language Dagbani:

(36) Dagbani (Lord 1993: 128)
m zang m suu nmaai nïmdi
I took my knife cut.PFV meat
'I cut the meat with my knife'

In this Dagbani example, *zang* is the equivalent of a preposition such as *with*. The knife does not have to be moved—it may already be in my hand.

11.5.2 Grammatical separation of C1 and C2 is enhanced

In a second group of examples, the grammatical separation of the two clauses is enhanced in some way.

(i) The subject of C2 is not ellipted

In this very common variant, the subject of the transitive verb in C2 is not omitted:

(37) I take time that normally would be devoted to memorizing laws and corollaries and I invest it in different types of experience.

(38) so sometimes we take three or four characters and we make them into one.

(39) Santa Barbara Corpus (Conv. 3):
PETE: . . You know,
. . the early man probably said . . . the same thing about the first domestic chicken.
MARILYN: . . (H) . . <VOX Can you imagine,
they're taking this beautiful jungle fowl,
and they're leaving it in a box.
That's gross VOX>.

The presence of a separate subject in each clause moves the construction further away from standard serialization, in which the two verbs share a single subject, and emphasizes the separateness of the two clauses.

(ii) The conjunction is other than *and*

The all-purpose "vanilla" coordinator *and* can be replaced by a different coordinator, generally *but*.

(40) And he could take the most complex story—he himself was a very complex man—but tell it in the most, not simplistic, but simple manner; very direct.

We saw that the presence of *and* is itself an obstacle to a serial verb interpretation of the *take NP and* construction, since it seems to constitute a clause boundary. Above, we defended the inclusion of *and* on the grounds that it was obligatory in English coordination, and therefore did not alone disqualify hendiadic constructions from being viewed as serializing. However, *but*, with its contrastive force, throws the second clause into relief, and seems to move the construction a step away from canonical serialization.

(iii) There are multiple C1s ('take' clauses)

Examples of this type are given in (41)–(43):

(41) it's nice to hear. I just wonder what would happen if we were to take a choir and take somebody from all of these different parts of the country and put them in together what would we get? What kind of a sound would we get?

(42) We want to be able to go into the schools and take a penis model and take a condom and show these children how to use it.

(43) My hope is that we can take that energy and take that anger and take that that unrest that people are feeling and do something with it to really make some differences in public policy.

There is obviously a strong pragmatic effect that is achieved by declaring the objects of the verbs first with the *take* clause and suspending announcement of the principal verb (*show* in (42), *do* in (43)). The direct objects are dramatically charged (*penis*, *condom* in (42), *energy*, *anger*, *unrest* in (43)). The rhetorical figure is the *period* or *periodic sentence*, characterized by an accumulation of ideas that cannot be fully understood until the last verb. This figure is sometimes held to be more natural in Greek and Latin because of the grammatical possibility in those languages of the sentence-final verb, which permits the climax of the sentence to be postponed until the very last. It is interesting, however, that the *take NP and* construction, by deferring the lexical verb until the last period, enables the classical figure to be replicated in English.

11.5.3 Non-canonical instances of C2

The last group presents the greatest challenge to the idea of *take NP and* as a construction. In this set of examples, the integrity of C2 is itself placed into question, and C1, the *take* clause, seems to stand alone.

(i) There are multiple C2s

In this rather rare variant of the schema, the verb *take* is resumed not by one but by several transitive verbs in follow-up clauses.

(44) . . . when Gerard was sitting there expressionless, sometimes even seeming to doze, in fact he was listening to everyone, he was ingesting their technique, and finally he took their technique, transformed it, and surpassed it.

At first sight it might seem that this type is simply the mirror image of the type in which there are multiple C2s instead of multiple *take* clauses. However, with multiple *take* clauses, the last *take* clause is still followed immediately by a C2, whose verb resumes the most recent instance of *take* (i.e., in (42) the verb *use* resumes the *take* of *take a condom* rather than the *take* of *take a penis model*). In (44), on the other hand, the series of coordinated clauses is unbounded, and each subsequent clause moves its verb further away from the *take* in whose scope it lies.

(ii) C2 is grammatically and/or semantically disconnected from C1

In this variant of the canonical schema, there is still a recognizable follow-up clause, but no element of C2 has a grammatical connection with any element of C1. The relationship between the two clauses is purely pragmatic. The "implied instrumental" type represented in examples (32)–(35) could be included here; however, in those examples there is still a consistent semantic relationship between C1 and C2 that is not found in the next set.

(45) We can't take every problem of society and say that city government and state government has done that because of minorities.

(46) It's probably a bit unfair on a new lad to sort of er er make you know sort er of take him take his circumstances and generalize it.

(47) Nevertheless, from the outset the First Lady resisted full disclosure of relevant documents, arguing that 'the press will take them and twist it and put it in the worst possible light and it will give our enemies ammunition'.

Even here it could be argued that *that* in (45) and *it* in (46) and (47) are "sloppy" anaphors and still have a loose syntactic relationship with the direct object of *take* in the preceding clause. In the next group not even this much can be claimed.

(iii) There is no clear right-hand boundary that would signal a natural syntactic unit, i.e., *there is no identifiable C2*

This set comprises instances where any direct relationship between C1 and C2 is completely lacking. Since one can always project an inferential relationship if one tries, the difference between them and the examples in the preceding group is sometimes rather subjective. (In these examples, which are necessarily rather long, the verb *take* has been highlighted with bold face.)

(48) This year they reduced the reversion, at our request, from 2 percent of our personnel budget to 1 percent and they directed that we spend that money on technology and libraries, so we are going to **take** some of that money that is directed to be spent on technology and issue another request for proposals from faculty this time to focus on putting your course on the internet or developing a course for the internet if you don't want to put on the internet one that you've got already. (CSPAE)

(49) And what we can do with it is to look at how we can **take** what we've done thus far, what we've learned from those experiences and how we can help make this the best one possible. (CSPAE)

(50) QUESTIONER: And if I could follow up—the Croat Muslim federation has said that they need at least 50 percent of the land in Bosnia?
SENIOR ADMINISTRATION OFFICIAL: Well, if you **take** the parameters that have been on the table in Geneva, and you add up the Muslim plus the Croat entities, you get just about 51 percent. (CSPAE)

(51) I think this conference actually will be very interesting. I really do. You know, if you **take** day to day events, it's not what this conference is about. But if you **take** the things that are really on these ministers' minds and I know that we are thinking about all the time, and they really are the framework within which—both substantive and political—framework within which all of these decisions they make are made, that's what this conference is about. (CSPAE)

(52) Having said that, let me try to suggest what some of these new windows are, what's the context here. Well let me **take** some ideas, some of which seem to be in opposition to each other but just by way of illustrating some things. Standards have always been here. In the paper world we've always had standards for how we do business. Fill out this form, pass it along to this next office, get it out there, pass it along, and so forth. We've had standards. They're very rigid. (CSPAE)

(53) That is, they are **taking** bits and pieces of federal program and they're twisting like pretzels and they're putting things together. And they have had a very difficult time doing it, but they're already doing it. (CSPAE)

(54) Those of you who are PIs and who have pushed support personnel into research grants have typically **taken** the graduate student and put part of that graduate student support in that way, along with whatever else you've been able to put together with the package. (CSPAE)

(55) I think that they're, they're operating on a more, mean, I know that they're causing massive problems in society up here, but I don't really think that, that it's, it's in our power to **take** these people from a sovereign state and say you don't deserve to live (SWITCHBOARD)

(56) Santa Barbara Corpus (Conv. 9)

NATHAN: You got the two,
and you **take** the square [root of two],
KATHY: [@@]
NATHAN: and you get the negative [2two2],
KATHY: [2@2]
NATHAN: which you **take** [3.. the square,
KATHY: [3@@@@ (H)3]
NATHAN: and it comes3] to two,
KATHY: @@@ (H)
I'm sorry,
(H)
NATHAN: (Hx) … So.
… let's talk about this slow=ly=,

(57) Actually, if you look, what I did is I **took** the last five years of data by division and by rank and I saw also where the lowest percentages are for women and it seems to be that at the instructor and assistant professor level proportions are kind of high but when it comes to the associate and full professor level this is bringing people in in the last five years we have the lowest percentages. (CSPAE)

The verb *take* is here in a grammatical clause of its own, without any discernible follow-up. We must therefore view it as having a semantic content that is somewhat different from that in the canonical schema, where *take* is a purely abstract grammatical formative serving to anticipate an upcoming transitive verb. We might in the examples in group 11.5.3(iii) paraphrase *take* with some verb like *consider*, *start with*, or *look at*, or sometimes even *take possession of*, serving to introduce a topic for an undefined stretch of discourse. There would be no assumption that this topic should be explicitly mentioned in the following discourse, only that it anticipate a generally relevant theme. In (57), the speaker identifies the domain of the upcoming discourse as "the last five years of data by division and rank". There is no expectation that these five years will themselves figure explicitly in the discourse, but the audience understands that the various items in the discourse are framed by this period of time.

Obviously the use of *take* in the common sense of 'take as an example' also belongs here. Instances of this are very numerous; see, e.g., (58)–(59):

(58) Take a country like Britain where the population is absolutely stable. It's been declining slightly but it's you know it's a no change population.

(59) ...if I may press you for your thoughts on them. Erm take the word 'hell' for instance. Erm I was glad to see that under 'hell' there was a helpful list of phrases erm involving the word 'hell'. Take these two phrases 'there'll be hell to pay' and 'they went at it hell for leather'. One of those is in one of them isn't. Now if a computer tells me that one of those ought to be in and one of them ...

This is also, as I shall show (11.7.3 below), the oldest manifestation of the discourse use of *take*.

11.6 *TAKE NP AND* FROM A TYPOLOGICAL VIEWPOINT

The use of the verb *take* as the first component of a serial verb construction is so pervasive and widespread in the world's languages that we are justified in tentatively comparing English *take NP and* with the canonical schema for serial verbs.[9] We will consider the use of *take NP and* from the perspective of two typological features, *transitivity* and *morphology*.

11.6.1 Transitivity

The correlation between high transitivity and serialization with *take* is noted by Carol Lord (1993). In the West African language Nupe, a verb translatable as *take* has a causative function, converting for example 'be in a place' into 'put something into a place', as in (60):

(60) Nupe (Lord 1993: 126)
a. sàlàmí là ébi
Salami took knife
'Salami took the knife'
b. ébi ta ésàkí o
knife be table on
'the knife is on the table'
c. sàlàmí là ébi ta ésàkí o
Salami took knife be table on
'Salami put the knife on the table'

[9] This is not to suggest that in serial verb languages the serial verb construction in each case does not need to be studied from a detailed discourse perspective. Yusuf (1986), in which it is shown that Yoruba serialization is in actual discourse a rudimentary and lexically restricted construction, is an important example of the necessity to do just this.

We saw earlier that one feature of the Chinese *bà* construction, formerly 'take' but now exclusively grammaticalized as a "pretransitivising" morpheme (Chao 1968: 342), is *completion* ("telicity" in the Hopper and Thompson transitivity metric), as previously illustrated in example (8), repeated here as (61):

(61) Chinese (Sun 1996)
a. ta he le tang le, keshi mei he-wan
3SG drink ASP soup ASP but NEG drink-finish
'He has eaten the soup but did not finish it'
b. ta ba tang he le
3SG BA/take soup drink ASP
'He has eaten the soup (and finished it)'

A similar semantic force is associated with *take* in Hindi, where there is a coverb *lenaa* (past tense *liyaa*) the use of which connotes a successfully completed action.

(62) Hindi (Hook 1974: 166; morphemic analysis here by PJH)
a. maiN ne aap ko das baje fon kiyaa
I ERG you DAT 10 o'clock phone made
'I rang you [dialed your number] at 10 o'clock'
b. maiN ne aap ko das baje fon kar liyaa
I ERG you DAT 10 o'clock phone make took
'I called you [and got through] at 10 o'clock'

Grammaticalization of *take* as a valency raising morpheme is not at all uncommon, and it is not surprising to find that the *take NP and* construction has the hallmarks of *high transitivity* as defined in Hopper and Thompson (1980). It is easy to construct English sentences with *take* in which an increase in transitivity is evident and in which a transitivity threshold decides the difference between the possible use vs. non-use of *take.* For example, *take* may not be used if the verb in C2 would be effective. Effective verbs bring about their object; because they do not change an already identified object, they are less transitive. Affective verbs, which change their object, are more transitive (cf. Hopper 1986). Thus lighting a candle (an already present object) is more transitive, whereas lighting a fire (the object comes about through the act of lighting) is less transitive.[10] Only the first can participate in the *take NP and* construction:

(63¢) a. I took a candle and lit it (= I lit a candle)
b. *I took a fire and lit it (= I lit a fire)

If this perspective is correct, we might expect that verb-serializing languages would show evidence of a grammaticalization cline such as in Figure 11.1.

In her article on transitivity in serial verbs in West African languages, Carol Lord (1982) notes precisely this kind of variation. In Ga (Ghana), there is a co-verb *kɛ* that translates as 'take' and which is in construction with transitive verbs like *ngmesi* 'put down'.

[10] On effective verbs ("resultative" verbs), see Jespersen (1933: 109).

FIGURE 11.1. Cline of grammaticalization of effective verbs

(64) Ga
a. e kɛ wolo ngmesi
she take book put.down
'She put the book down'
b. *e kɛ wɔlɔ ngme
she take egg lay
'She laid an egg'

But if the second verb is effective, *kɛ* may not be used, and sentences like (64b) are ungrammatical.

Similarly in Akan, where there is a co-verb *de* 'take' (as in 65a) that cannot be used with effective objects in sentences like (65b):

(65) Akan
a. Kofi de nwoma no a-ba
Kofi take book that PF-come
'Kofi brought that book'
b. w-a-kyerwɛ me nhoma
he-PF-write me letter
'He wrote me a letter'

There are also languages in the same geographical region in which the *take* verb can co-occur with both affective and effective verbs, e.g., Idoma as in (66):

(66) Idoma
a. o l uwa nu
she take them drive-away
'She drove them away'
b. o l ɔyi ma
she take child bear
'She bore a child'

The presence of still prior points on the continuum is suggested by Dagbani, where the *take* verb allows only direct objects that are movable; the equivalent of *take truck paint* for 'paint the truck' is possible, but not *take room paint* for 'paint the room'.

Furthermore, *take* in the hendiadic construction has a stricter expectation of definiteness/referentiality in the direct object than *take* in non-hendiadic sequences, and this too is a by-product of high transitivity. The statistics given in Table 11.2 on a sample of 90 hendiadic (monopredicative) examples and 106 polypredicative examples

TABLE 11.2. Definiteness in *take NP and*

	Proportion of definite-referential objects
monopredicative "take NP and V"	81% (total n = 90)
polypredicative "take NP and V"	59% (total n = 106)

show a significant bias toward definite-referential direct objects when *take NP and* is monopredicative.

Not only in semantics, but in discourse function also, the *take NP and* construction conforms to the expected discourse picture of high transitivity. For *take NP and* has an affinity with discourse foregrounding and the identification of key points in a developing discourse. It highlights the places that the speaker wants to be memorable and worthy of attention.

11.6.2 Morphology

The phenomenon of verb serialization, once thought to be limited to a small group of African languages, is now known to exist in almost every major language area. Earlier researches attempted to reconcile the evident presence of two distinct verbs in a single predicate with a syntactic theory that provided for only one verb in a predicate, either by postulating a deep structure with two clauses, or by analyzing one of the verbs as an auxiliary. An important step forward in understanding the nature of serial verbs was the recognition of a historical dimension visible through the lens of linguistic typology and the study of grammaticalization (Lord 1993; Aikhenvald 1999). Serialization could now be seen as part of a more general historical process that included secondary verb phenomena such as auxiliation (Kuteva 2001), verb compounding (Aikhenvald 1999), and the "vector verbs" of Indo-Aryan languages (Hook 1974; Aikhenvald 1999). In her work on the typology of serialization, Aikhenvald makes the observation that verb-serializing languages are generally poor in morphology. The reason for this is clearly that morphology, especially verb morphology, frequently works to delimit clauses, and the merging of clauses is essential to the emergence of true serialization. Even in languages with some morphology, it is commonly found that the full range of morphological contrasts is not available in serialized verb constructions. Can it be that the emergence of serialization goes hand in hand with the loss of morphology?

English might be seen as a test case for this possibility, since it has lost its morphology in historical times. If serialization comes about in the context of morphological simplification, we might profitably look for signs of incipient serialization in English. Such signs would not be the sudden appearance of serialization, but would manifest themselves in a more subtle way, perhaps as a loss of inflectional contrast in hendiadic sequences. We noted that this is precisely what has happened in the case of *try and*, a construction in which the other forms of the lemma *TRY* (*tries*, *trying*, *tried*) are completely absent.

TABLE 11.3a. Incidence of forms of the lemma *TAKE* in *take NP and* and of all forms of *TAKE* in Cobuild

	TAKE in *take NP and*	All instances of *TAKE*
take	61% (n=55)	42%
taking	15% (n=13)	14%
took	12% (n=11)	20%
takes	9% (n=8)	8%
taken	3% (n=3)	16%

Corpus studies show that *take NP and* is not categorically limited in the same way as *try and*. Unlike *TRY*, all members of the lemma *TAKE* are found:

(67) a. Nothing wrong with that but that's not the same as taking the two models and screwing them together
b. She takes this analysis and applies it to the most controversial section of her book
c. But while Ford was catapulted upwards in *Star Wars* and the Indiana Jones epics, Hanks has taken innocent little stories and turned them into successes
d. The station took familiar songs and changed their words

Yet when we examine the distribution of the different forms of the lemma, a significant fact emerges. In the hendiadic *take NP and* construction the uninflected form *take* dominates other forms of the lemma numerically to a much greater degree than when the sequence is synthetonic. Table 11.3a shows a comparison of all forms of *TAKE* in the *take NP and* construction and at large (data from Cobuild only).

Table 11.3b summarizes these figures by consolidating all inflected forms of *TAKE*. Among the forms of *TAKE* in the Cobuild Corpus as a whole, the uninflected form *take* has the largest share, but with 42 percent not an absolute majority. In the hendiadic construction the form *take* accounts for almost two-thirds of all instances of *TAKE* (61 percent), and the share of inflected forms drops from 58 to 39 percent. If the uninflected *take* is the clear winner, the losers are what might be called the "narrative" forms—the past tense *took* and the perfect/passive participle *taken*, suggesting that a narrowing of genre may be a contributing factor in the redistribution: as I will show below, the *take NP and* construction has an affinity with the building of arguments rather than with reports of the past. Whatever the reason for the shift, the increased

TABLE 11.3b. Inflected versus uninflected (base) forms of *TAKE* (summary of Table 11.3a)

	take NP and	All instances of *TAKE*
Uninflected	61%	42%
Inflected	39%	58%

discourse prominence of the uninflected form is consistent with the general picture of emergent serialization in *take NP and* that is being proposed here.[11]

11.7 THE GREATER *TAKE NP AND* CONSTRUCTION

Previously we saw that although in the main there is plenty of reason to identify a construction conforming to the canonical schema, there are different degrees of closeness between the two clauses. They range from the canonical schema itself, in which *take* and its partner verb constitute a single functional entity, to what amounts to a complete absence of grammatical, semantic, or pragmatic connectedness between the two clauses, where only a special use of the verb *take* testifies to any kind of link with the canonical schema. The tighter of the two poles, represented by the canonical schema, is itself close to the monopredicative prototype offered by the serial verb typology, in which a single grammatical clause accommodates both verbs. The different deviations from the prototype sometimes allow themselves to be placed at discrete points along the continuum and sometimes not; obviously much depends on the criteria chosen. Tentatively, a continuum can be posited, according to the degree to which C1 (the *take* clause) and C2 (the follow-up clause) are conceptually and functionally integrated. The points on this continuum are shown in Figure 11.2. Figure 11.3 shows the continuum itself and its relationship to the use of *take* as a serial verb in serializing languages.

There may be some further hierarchization of the constituents within column 6 of Figure 11.2, though this is not obviously the case. Theoretically, the continuum reaches beyond English into the typological area of full serialization, though it should be stressed that there is no reason to think that the English *take NP and* construction itself will ever move in this further direction. Nor do we need to postulate that the non-canonical variants will eventually be shed, leaving a pristine point 7 as the only relic of a diachronic process. In fact, to separate the non-canonical parts of the continuum (i.e., points 1–6) from the canonical schema (point 7) would be quite artificial. There is no place at which the continuum can be naturally bisected, that is, no place on the continuum where two adjacent clauses could be said to definitely constitute a conceptual unit as opposed to some other place where they are definitely conceptually distinct. Rather, when we speak of the *take NP and* "construction" we should have in mind the entire continuum, not just the extreme end that gives it its name. Metaphorically speaking,

[11] The correlation between loss of inflections and emergent serialization also sheds light on the curious fact that *take NP and* does not appear to have a counterpart in French and German, inflected languages that otherwise share many idioms and constructions with English. The observation by Schiller (1990) and Whitman (this volume) that serialization is rare in strictly verb-final languages is, if true, thus no doubt susceptible of a natural diachronic explanation, the common factor being loss of the inflectional suffixes characteristic of verb-final languages. A historical connection between the erosion of inflection and the emergence of free or medial-verb word order has often been suggested.

Autonomous *take* (i.e., no C2)	Relationship of C2 to C1 is inferential but not grammatical (loose inference)	No C2 anaphor, but object of C1 is an instrument by means of which the action of C2 is accomplished (tight inference)	There are multiple C2s	*but* replaces *and* in C2	-There is an overt subject in C2 -There are multiple C1s -C2 anaphor is in a subordinate clause -C2 anaphor is in a possessive form -C2 anaphor is a prepositional complement	Canonical *take NP and* construction
1	2	3	4	5	6	7
Less integrated ←						→ More integrated

FIGURE 11.2. Points on the continuum of integration of *take NP and*

the *take* clause exemplified by point 1 on the continuum "pulls" a stretch of adjacent discourse into its scope, sometimes merging with it grammatically in a construction that is superficially biclausal but conceptually unified.

11.7.1 Implications of this analysis

We may hypothesize that there is something more general happening here. Many grammatical constructions are *biclausal*, that is, they consist of two identifiable clauses that partner one another and coexist in a tighter or looser formation, presenting some kind of unitary discourse function. Examples would be the pseudocleft (also known as the *wh*-cleft), as in (68), as well as its various congeners such as *the point is*, *the thing is*, *one of the ways . . . is*, etc. (Hopper to appear); the *it*-cleft, as in (69), and the *if*-conditional, as in (70):

(68¢) What worries us most is the long-term effects of this policy

(69¢) It is the long-term effects of this policy that worry us most

(70¢) If this policy is continued, the long-term effects will be disastrous

Point 1 (*take* in C1 but no C2)	Points 2—6	Point 7 (Integration of C1 and C2 in the *take NP and* construction)	Full serialization
← English →			Serializing languages
Less integrated ←		→ More integrated	

FIGURE 11.3. Relationship of the continuum of integration to full serialization

Discourse studies of such constructions generally start with the assumption of a unified construction in which examples like (68)–(70) serve as prototypes. There is already evidence that the first of these, the pseudocleft, works in a similar way to what has been described for *take NP and* in the present paper: a *wh*-fragment annexes a certain stretch of subsequent discourse, which may or may not be a grammatical predicate (Hopper 2002, to appear). But the type with a grammatical predicate assumes some kind of linguistic priority, and eventually comes to be grammaticalized and recognized as the prototype. Sanderman's recent work on the *if*-conditional (Sanderman 2004) suggests a similar discourse situation here too: her study of the distribution of *if*-clauses shows that as many as 36 percent of protasis *if*-clauses are not in fact followed by an apodosis (corpus examples include *if we assume that*, *if you don't mind*, *if you say so*, etc.).

11.7.2 Older explanations of biclausal sentences

The pre-generative literature on historical syntax contains many references to the textual circumstances under which complex sentences emerge. Question-and-answer pairs, for example, have often been postulated as the source of certain types of biclausal sentence. For Hittite co-relative clauses, Held remarks that "[the indefinite pronouns] *kuis/kuit* may be relatives or interrogatives, depending on whether the two clauses are to be taken together or individually" (Held 1957: 40).

This kind of explanation has long been standard in the case of Germanic initial verb conditionals, such as English *had we noticed the leak, we would have called a plumber.* An originally indicative *had* of an independent interrogative was replaced with the subjunctive when (or as) the sequence became grammaticalized as a biclausal conditional (Einenkel 1916: 43). There are comparable constructions, with the same explanation, in German and Nordic languages. Several earlier authors point out the close parallelism between conditionals and concessives, both characterized by an initial auxiliary (cf. German ***hatte*** *er ihr doch Blumen gebracht* "after all, he **had** brought flowers for her"). Gardiner (1932: 228–229) notes for Middle Egyptian and Coptic a change from a corroborative question *'in 'iw* "is it (the case that)?" to the protasis of a conditional. See, too, Jespersen 1924: 305. In a similar (post-generative) vein, Herring (1991: 273) describes for Tamil a situation in which biclausal sentences alternate with equivalent question-and-answer sequences:

(71) avan ên / inkê illai nnâ, avan ûrukku pônân
he why RISING INTONATION here NEG SUBD he town.DAT go.PAST
'Why isn't he here? He went to the village'

(72) avan inkê illai ênnâ, avan ûrukku pônân
he here NEG because he town.DAT go.PAST
'He isn't here because he went to the village'

Haiman (1979) has noted the strong affinity between the protasis-apodosis of conditionals and topic-comment structure. The protasis-apodosis fixture of conditionals is

therefore perhaps better seen as another instance of the 'capturing' by one clause of a piece of subsequent discourse and their combined grammaticalization as a biclausal construction. In such cases there is evidence that topic-creating questions (perhaps rhetorical ones) have evolved into the protasis of a complex syntactic construction, for which the discourse mechanism is through a particular utterance-type that has come to be understood as anticipating, or *projecting*, an upcoming stretch of discourse.

11.7.3 Historical background of the *take NP and* construction

The account of the *take NP and* construction offered here suggests that the earliest forms of the construction should be the less integrated ones to the left of the continuum of integration. Is there textual evidence pointing in this direction?

The verb *take* is a Nordic form, borrowed into English during the later Old English period. For several centuries it competed with its native equivalent *niman*. The two verbs coexisted until the Middle English period, when quite rapidly *niman* was ousted by *tacan* (Rynell 1948). The Old Norse verb *taka* was already available in contexts that point to grammaticalization as an aspectual marker, as in ***tók** hann þá ok herjaði bæði útanlands ok innanlands* 'He set about and laid waste both inland and abroad' (cited in Norde 2003). It is not clear whether this use of *take*, which is widespread in Europe, results from syntactic diffusion as suggested by Coseriu (1966), or is an independent development. Certainly by the Late Middle English period a transitive use of the imperative *take* existed that served to present a staged series of instructions organized around a single entity. An early such use of *take* is (73), from the year 1393.

(73) Take a metal plate or else a board that has been smoothly shaved with a level and polished even, whose whole diameter, when it has been rounded by means of a compass, shall contain 72 large inches or else 6 feet of measure, and which has not warped or bent. The edge of the circumference shall be bound with a strip of iron in the manner of a cartwheel. The board itself may, if you wish, be varnished or glued over with parchment for permanence. Take then a circle of metal 2 inches in width so that the whole diameter within this circle contains 68 inches or 5 feet 8 inches and carefully let this circle be nailed around the circumference of this board or else make this circle out of glued parchment.[12]

This transitive imperative use of *take*, which is the presumed ancestor of the *take NP and* construction, existed already in Old Norse and appears to have an unbroken

[12] The example, which is from the Helsinki Corpus, is from a medieval text *The Equatorie of the Planetis*. The translation into Present-Day English is my own. The Middle English original is as follows:

tak therfore a plate of metal or elles a bord that be smothe shaue by leuel & euene polised of which whan it is rownd by compas the hole diametre shal contene .72. large enches or elles .6. fote of mesure, the whiche rownde bord for it shal nat werpe ne krooke; the egge of the circumference shal be bownde with a plate of yren in maner of a karte whel. In this bord yif the likith may be vernissed or elles glewed with perchemyn for honestyte. Tak thanne a cercle of metal that be .2. enche of brede & þæt the hole dyametre with in this cercle shal contene the forseide .68. enches or .5 fote & .8. enches & subtili lat this cercle be nayled vp on the circumference of this bord or ellis mak this cercle of glewed perchemyn.

history in English from the time of the first uses of the borrowed verb. The original transitive sense of *take*, as suggested by Gothic *tekan* 'to touch' and the Latin cognate *tangere*,[13] was a quite concrete one, and most early uses of *tacan* involve capturing, handling, removing, or otherwise taking in a quite physical sense. Its use to introduce a discourse theme was rare to begin with, but was often associated with procedural genres (including, importantly, culinary and medical prescriptions). The earliest example of this usage cited in the *Oxford English Dictionary*, given in (74), dates from *c.*1200:

(74) Tacc nu þiss streon þatt tuss wass sibb Wiþ preostess & wiþ kingess
'Take now this offspring that thus was kin to priests and kings'[14]

But the real burgeoning of the transitive discourse use of *take* took place in seventeenth-century scientific discourse. Careful description of experiments required a sequenced account in which the steps were linked by a common referent. This referent, once it had been introduced by *take*, could from that point on serve as a common focus for the account, as in (75)–(76):

(75) And in prosecution of this Experiment, having taken the filings of Iron and Steel, and with the point of a Knife cast them through the flame of a Candle, I observed where some conspicuous shining Particles fell, and looking on them with my [Microscope], I found them to be nothing else but such round Globules. (1665)[15]

(76) Fifthly, Having taken some Amber, and warily distill'd it, not with Sand, or powder'd Brick, or some such additament as Chymists are wont to use, for fear it should boyl over or break their Vessels; but by it self, that I might have an unmixed Caput mortuum; Having made this Distillation, I say, and continued it till it had afforded a good proportion of phlegm, Spirit, Volatile Salt, and Oyl, the Retort was warily broken . . . (1675–6)[16]

The extension of the locution from the simple uninflected imperative (*tacc*, *take*) to fully inflected forms of the verb like *having taken* suggests the emergence of a more integrated, syntacticized structure having its beginning in the Renaissance and thriving in the context of seventeenth-century science. Ultimately, then, the explanation of how the uses of *take* became syntactically elaborated must be sought in broader cultural currents, including printing, the decline of orality, and the rise of a quite conscious scientific rhetoric.

11.8 CONCLUSIONS

I hope in this paper to have shown, through an examination of the *take NP and* construction, how a close study of the distribution of a construction in discourse leads

[13] The correspondence Gothic *t* ~ Latin *t* is irregular, but the semantic, morphological, and other phonological similarities are striking.
[14] *Oxford English Dictionary*, sub *take (v)*.
[15] Robert Hooke, "Of the fiery sparks struck from a flint or steel". In his *Micrographia*. Helsinki Corpus.
[16] Robert Boyle, *Electricity and Magnetism*. Helsinki Corpus.

to a view of grammar as something fluid and unstable, that is, as emergent from, and inseparable from, its discourse environment (Hopper 1987). Verb serialization, in this view, would then be one part of an entire range of possible uses of the first verb, some of which have become grammaticalized, and which are in turn part of a general process whereby a word with a wide range of meanings *projects* subsequent stretches of discourse, 'captures' them, and hauls them into its scope. Verb serialization is thus, to the extent that it is not a borrowed feature, an *emergent* process.

This position should be sharply distinguished from one that starts at the opposite assumption, that a language is a set of fixed constructions each with a mental prototype defining a central instance and existing independently of discourse. The picture that suggests itself is much more complex. The various forms that constitute a construction are a family-resemblance array (Hopper 2001). Like all constructions, *take NP and* is open and unbounded (Hopper to appear). In the case of *take NP and*, the extreme points of this array are at one end a special use of the verb *take* to introduce a new discourse theme, and at the other a merger of a *take* clause and a following transitive clause. In between are different kinds of inferential relationships between the *take* clause and the subsequent discourse, as well as partial and overlapping grammatical similarities. And the different points on the conceptual network that is the *take NP and* construction retain their structural and functional interconnectedness.

11.8.1 Projection

The grammatical picture of the *take NP and* construction as presented here fits in well with recent studies of *projection* in conversation analysis.

In its standard sense, projection refers to a speaker's anticipation of the completion of a conversational turn. As such, it is limited to a rather brief horizon of expectation, specifically the end of the current conversational turn. Thus Ford (2001: 55) characterizes projection, or projectability, in the following way: turns "have structures the courses of which can be roughly predicted before a turn is completed". Through projectability, "a recipient can predict at what structural, prosodic, and pragmatic (action embodying) place in a turn's unfolding the turn is likely to end". But the *take NP and* construction of English is not found in short turns. On the contrary, it functions in part as a tactic for holding the floor past the next turn. It is characteristic of what are called *expanded turns*. This is why, as the above examples show, the construction is common in the "consultative style" (Joos 1962), those genres that are home to oral monologic explanation and reasoned argument, but rare in both written texts and in intimate conversation. The *take NP and* construction is antagonistic both to the quick back-and-forth of casual conversation and to writing, where the exigencies of speech (improvised composition of long turns, floor-holding) are absent. However, there is a role for projectability in the wider sense in which speakers (and listeners) have expectations about the "direction" of the discourse, the "gist" of what is about to be said, and pick up clues about it that are coded in words and constructions routinized for that purpose (Auer 2005). The bleached lexical verb *take* and its immediate direct

object is one of these *projectors*, as we might call them. The semantic meaning of *take* is especially suited to its function as a projector in that it has a deictic dimension away from the speaker (in contrast to *bring*), and thus points metaphorically to the as yet unrealized discourse.

The stretch of discourse projected by such forms can be said to lie in the *scope* of the projector. Scope in this sense varies in extent. It can be strictly local or it can fade away to generality, but clearly those forms that are adjacent to the projector are better candidates for grammaticalization as a biclausal syntactic construction.

11.8.2 Emergent Grammar

The view of grammatical structure presented here is that of Emergent Grammar (Hopper 1987). Grammatical structure is "a result of language use in context" (Laury 1997: 3), or a set of "routinized schemas and patterns, generalized from the structures which most frequently emerge in the fulfillment of speakers' communicative goals" (Englebretson 2003: 89) and is therefore in a constant state of adaptation to the pressures of discourse. Emergent Grammar is understood as epiphenomenal and derived from usage. Because the goals of speech are interactive ones, speech proceeds not by translating mental ideas into physical language, but by adapting previous utterances to the current needs of speakers and interlocutors, *including anticipated responses*. Its sources are not mental structures but previous utterances, which serve as rough models for current utterances. Repetitions of previous utterances cannot be identical to their source. They are not—indeed, cannot be—exact replicas, but only approximations, and so the models for subsequent utterances are always changed versions of the previous ones. Normally these changes are slight: for instance, a lexical item that extends a class of lexical items in a construction, or a subtle blend of two constructions, or some other kind of relaxation of constraints. It should be stressed that Emergent Grammar *is* a kind of grammar. It is not, as one linguist has simplistically alleged, a "no grammar" position. Grammatical forms do emerge out of discourse events; but this emergence is not a historical *fait accompli*, but an ongoing process.

The visible results of emergent structure at any one time are small-scale. Studies of Emergent Grammar are more likely to include the use of specific lexical items, such as *take*, than of more general and abstractly conceived constructions. It would be a mistake to restrict the notion of grammar to the highly visible rules of traditional descriptions (Passive, *it*-Cleft, etc.), whose major rules have worked themselves out over longer stretches of time. Yet it must be stressed that even these more salient kinds of rules do not in spoken discourse possess the hardness and invariance that is often uncritically attributed to them. Constructions that may appear to be quite stable and robust when idealized out of their discourse contexts invariably turn out, on closer inspection, to be labile and ragged and susceptible of adaptation to the immediate pressures of usage. The *take NP and* construction meets a variety of open-ended

communicative needs, both social-interactive and cognitive, and therefore assumes a variety of open-ended, complexly related forms.

11.8.3 Protasis and apodosis

We can say, then, that the sources of verb serialization lie in a discourse use of a semantically bleached verb such as *take* that comes to be grammaticalized as a protasis, and a projected action that comes to be grammaticalized as an apodosis, and that the merger of the two into a single constituent is facilitated by the absence of a general coordinator (an *and* word) and by attrition of morphology.

The term protasis is normally associated with the first part of a conditional, the *if*-clause, and the term apodosis with the second part, the *then*-clause. But it may be worth noting that the restriction to the conditional construction does not reflect the original domain of these two terms. They refer rather to the first and second clauses of any two-clause combination, of which the conditional is only one type. Not only the *if*-conditional, but the first (i.e., subordinate) clause of any biclausal construction is a protasis: concessionals with *although*, temporals with *when*, *while*, etc., causals with *because* and *since*, and many other kinds of adverbial clause qualify as protases. So the range of protasis and apodosis goes beyond the specific example of the *if-then* conditional with which it is normally associated.

Yet the etymology of these terms reveals something even more profound. Protasis means 'stretching forth'. It refers to the reaching out toward some piece of discourse ahead of the speaker. And apodosis is a 'giving back,' a rejoinder, an *ant-wort* ('against-word'). An apodosis is a *response* to a protasis. So the serialized verb construction is ultimately the sedimentation (Hopper 1987) of a natural disposition to organize longer discourses as preparatory and currently focal material, that is, as protasis and apodosis in the original sense of these terms.

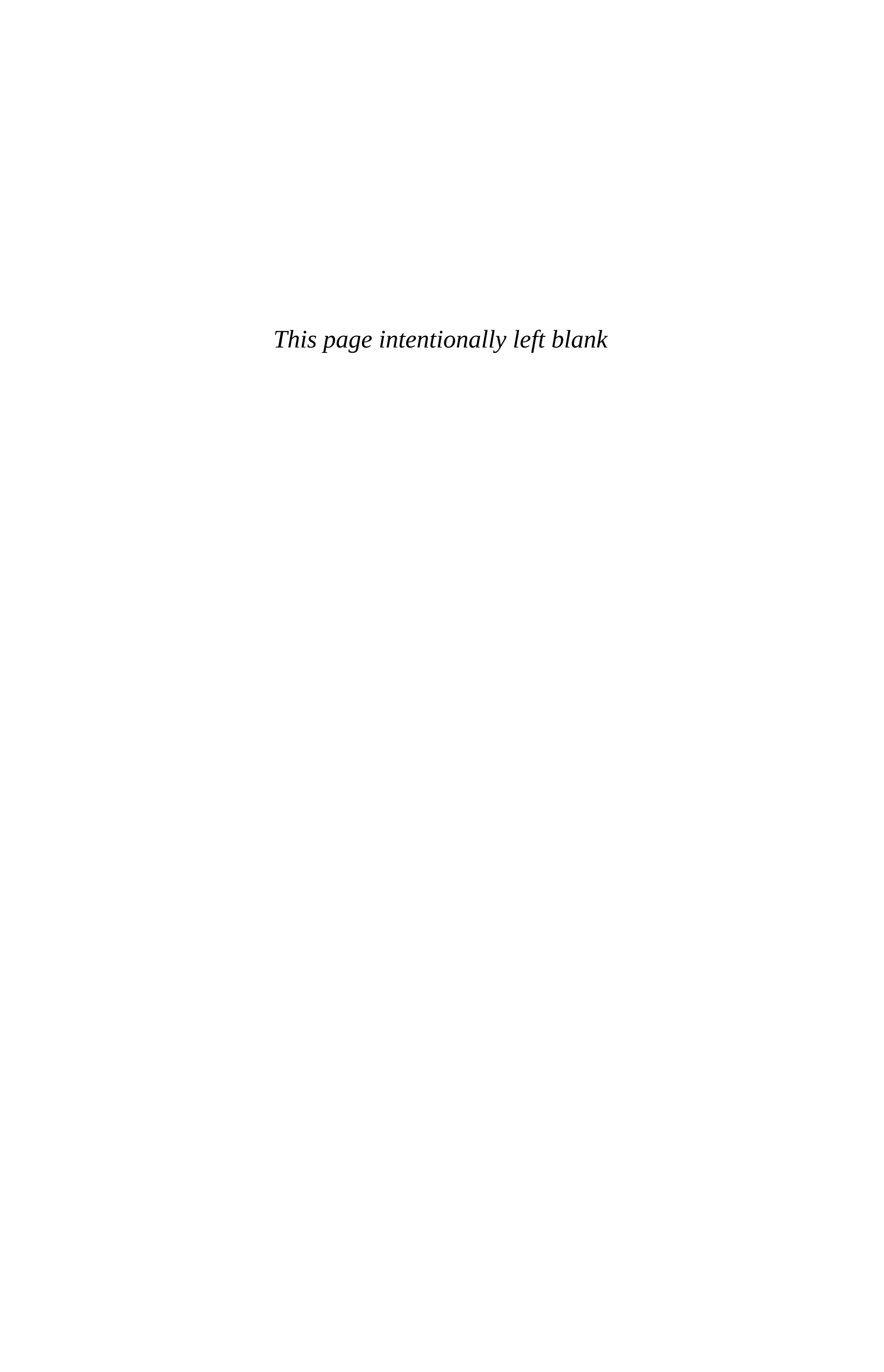
11.8.4 Temporality

Linguistics is beset with the paradox that although language is a real-time phenomenon, the methods and concepts of linguistics depend crucially on suspending the temporal dimension. Indeed, the rules and units devised by linguists are in the end simply what is left of language when the time dimension has been removed (Culler 1988; Hopper 1992; and Linell 2005, who attributes this move to the tyrannical hold of the written language on modern linguistics). One reason for this detemporalization of language is the theoretical stance that requires a *supervisory perspective* on utterances. In order to be amenable to linguistic analysis, utterances must already be fully present at the time of the analysis, and this presence requires synchronicity. Utterances must be seen not as unfolding in time but as complete and organic wholes that lend themselves to intricate hierarchical parsing. They must therefore be available fully and simultaneously from start to finish. But simultaneity is not compatible with the sort of

temporal linearity implicit in the account of the *take NP and* construction that is offered here. An utterance does not lead inexorably to a predictably structured apodosis, nor is an utterance an automatic response to a specific protasis. Rather, as Bakhtin insisted (1986), an utterance raises expectations of a response, expectations that can be met in full or in part or not at all. The protasis–apodosis relationship is linear-temporal rather than hierarchical-structural. It is based on an earlier versus a later temporal *phase* of the utterance rather than a simultaneous hierarchical assessment of the entire utterance (see Auer 2000).

It is clear that—like, ultimately, all grammatical constructions—*take NP and* has its origins in dialogic situations and is therefore irreducibly social in origin. It is deployed, not in order to facilitate a private monologue, but in order to manage a public discourse that involves a constant assessment of the interlocutor's comprehension and potential responses. Its representation, its "schema", emerges and is sustained and confirmed through its proven usefulness across numerous different interactive contexts. Linell (2005: 223) has noted that "we live in a dynamic, only partially shared and fragmentarily known, dialogically constituted world, in which relatively stable features (such as those of language and social representations) are emergent across series of communicative events". The serial verb construction *take NP and*, which is the local sedimentation of discourse protasis and apodosis, is thus in the end an outcome of basic conditions of language: temporality and dialogism. It is "emergent across [a] series of communicative events" in which a speaker announces a theme in summary form and elaborates it for the benefit of an interlocutor in a subsequent discourse. Through its demonstrated effectiveness it becomes a recurrent linguistic routine, and as its use spreads across genres, it compacts and grammaticalizes as the beginnings of a biclausal construction. It begins in rhetoric and ends in grammar.

PART VI
Conclusion



This page intentionally left blank

12

Universals and Diachrony: Some Observations

Johanna Nichols
University of California, Berkeley

12.1 INTRODUCTION

Over forty years ago, Joseph Greenberg took the first big step beyond a naive view of universals as things present in all languages, and presented the first explicit implicational and statistical universals, thereby laying the cornerstone for research on typology and universals. His subsequent work took a fairly teleological view of linguistic diachrony as following, revealing, and even implementing implicational universals. Work along these lines has been an off-again, on-again matter until recent years, when we have seen a good number of durable contributions to mainstream linguistics that have brought us to a sophisticated understanding of universals and their relationship to synchrony and diachrony. This book is evidence of that trend and gives several different and well-argued views of the connection between universals and diachrony: universals as structural implicational hierarchies available to all languages, not necessarily manifested in all but never reversed (Kiparsky); synchronic principles of economy (Haspelmath; Kuteva and Heine); more or less epiphenomenal consequences of ordinary diachronic processes (Harris) or constraints on diachronic processes (Blevins, Bybee; less directly Garrett, Albright); epiphenomenal consequences of trends in patterning (Hopper); a direct consequence of syntactic structure and constraints on movement or mapping (Whitman). A view not represented here is that universals are hard-wired into the human brain and diachrony is irrelevant to them. It should be noted that universals and rarities are two sides of the same coin, so, while the epiphenomena of language change described by Harris are rarities and not universals, they are still highly relevant to the understanding of universals.

A consensus view would appear to be that, rather than synchronic patterning always being the goal and driving force of language change, various synchronic patterns are the predictable consequences of diachronic processes which have their own logic

independent of the synchrony they produce. Thus, to a greater extent than Greenberg probably had in mind, synchronic structural patterns are epiphenomenal. But they are not entirely so. Economies of various kinds appear to be targets of change (as shown by Haspelmath, Garrett, and Albright), and there appear to be pure structural patterns that may be goals of change but are not its accidental results: Kiparsky's D hierarchy, word-final neutralization, stress-weight covariation; perhaps some of the word-order patterns, if they can be stated non-framework-internally; and the affinity between serialization and non-inflection that Hopper describes. This may be a consensus, but is this state of knowledge durable?

12.2 THE CURRENT STATE OF KNOWLEDGE

Whatever one's theoretical framework, there are some claims that are probably uncontroversial in all quarters: diachronic developments include some streamlining (and this affects synchronic structure); and synchronic structural regularities exist (for instance, structural types that are frequent worldwide and independent of family or area, such as the preference for SOV word order; implicational correlations such as those involving word order). Presumably uncontroversial desiderata for any analysis or theory include that it should be replicable and falsifiable, based on a systematic worldwide survey and not on cherry-picked positive cases, and expressed (or at least expressible) in framework-neutral terms and concepts. With this as background, let me address a remark or question or two to each of the chapters.

Kiparsky

How might the universal status of the D hierarchy, coda neutralization, and stress-weight covariation be falsified? Consider the D hierarchy, here convincingly shown to be available for exploitation to various grammatical ends. Further evidence is the fact that this same hierarchy acts in two different directions in the histories of genitive-accusative alternations in Slavic languages. In Russian, genitive-accusative syncretism replaces a former discrete accusative case beginning from the top of the hierarchy: personal pronouns have genitive-accusative syncretism from the very beginning of the written tradition; certain human nouns develop them early; ultimately, nearly all human and higher animate nouns develop them (in certain declension classes) (Klenin 1983). Also in Russian, a once-exceptionless rule requiring genitive instead of accusative on direct objects under negation is gradually weakening, resulting in accusatives under negation beginning at the lower end of the hierarchy (with indefinite and inanimate nouns, and only in those declension classes with distinct genitive and accusative forms) (Timberlake 1975, 2004: 321–327). Thus the genitive enters from the top and the accusative from the bottom.

However, there are parts of grammars where two discontinuous parts of the hierarchy are affected: e.g., "inalienable" possession, where cross-linguistically the "inalienables" usually comprise kin terms (the highest nouns in the hierarchy) and body parts (lower, following other human and all animate nouns) (Bickel and Nichols 2005; Nichols and Bickel 2005). Does this falsify the D hierarchy, or at least its strictly hierarchical nature? I assume the answer would be no, that this involves a different hierarchy, one based on frequency of use in possessive contexts (see Haspelmath), but when one hypothesis fails there will generally be something else that can be invoked. How do we falsify a claim that a hierarchy is available to all languages but not necessarily active in any language?

Harris

This is a convincing analysis but necessarily language-specific. Presumably a language with a different grammar to start with might arrive at endoclisis in some different way, using fewer steps (or more). A full evaluation of rarity might include not just an account of how a rare phenomenon did arise but also some measurement of different ways in which that outcome might arise, and their individual and joint probabilities.

Blevins

How might the claim that syllable structure is emergent be falsified? Also, this chapter cites a large number of positive cases of various epentheses, but it would also be desirable to do a general cross-linguistic survey and count the number of languages that do and do not use (or have and have not used) epenthesis.

Bybee

Structure Preservation, to judge from the examples here, operates during language use, including in adulthood, and during first-language acquisition. Or does it also constrain transmission or operate during acquisition? Also, how might one falsify the claim that synchronic patterns are emergent and the only true universals are diachronic processes?

Garrett

Does extension of non-alternating patterns necessarily involve whole paradigms? Or also subparts of paradigms? Can we expect to find cases where lesser alternations have replaced greater alternations, or where more frequent alternations have been extended at the expense of less frequent ones? All of these would also seem to involve extensions of existing paradigms (rather than targeting of simplicity *per se*) and to somehow reduce the complexity of a language. Also, are there enough languages

or language families for which we have comparative or historical information to be able to design a cross-linguistic sample testing whether or not leveling always involves extension of existing paradigms? It is my impression that a small but adequate sample could be designed, and this would seem to be a priority for understanding mechanisms of language change.

Albright

In the model proposed here, who is the learner—a child first-language learner? An adult (or at least non-child) fluent speaker acquiring new vocabulary? An adult second-language learner? Does the model handle equally well the kinds of changes that occur in normal transmission with minimal contact, normal transmission with contact, and non-normal (or at least what Dahl 2004a calls *suboptimal*) transmission? If so, it should be able to describe the simplifications that occur in creolization just as well as the changes that happen in ordinary transmission situations.

A priority for further research would be seeking or identifying test cases where Albright's model and Garrett's theory do not or cannot both apply, and establishing which wins out in the event of conflict.

Haspelmath

This chapter makes the very strong claim that *all* universal morphosyntactic asymmetries have frequency-based economic motivation, defining *universal* as "those that recur in language after language" (section 8.2). Recurrent patterns in minority language types are presumably not excluded just because they are in the minority; for instance, I assume there is an economy-based analysis of the typical zero ending of the absolutive case in ergative languages just as there is one for the nominative case in accusative languages, though ergative is a minority alignment type. But what of recurrent minority patterns that seem to reverse economy? An example is pairs of plain and semantic causative verbs (*fear: frighten, break (intrans.): break (trans.)*, etc.) in languages of the detransitivizing type, in which intransitives are often derived from transitives (which are simplex) rather than vice versa (Nichols, Peterson, and Barnes 2004). Most Indo-European languages are detransitivizing and in such pairs have a derived intransitive and an underived transitive, the clearest case being Russian with its many pairs such as *bespokoit'sja* (reflexive) 'be upset': *bespokoit'* 'upset, worry'. In corpus searches, the (longer) reflexive form is always much more frequent than the (shorter) non-reflexive. This pattern is systematic in languages of this (minority) type, so I would consider it a universal asymmetry and a systematic exception to the strong claim.

My more general questions here are: How is it decided what does and does not fall under "universal"? And how do speakers assess frequency, especially for words that fall in the middle of the frequency range?

Kuteva and Heine

I would have thought that, if cross-categorial harmony were important, there would be resistance to borrowing violations of it. Are the borrowings described here falsifications of the idea of cross-categorial harmony? Perhaps a theory of what is and is not likely to be borrowed would explain this. Also, I would have thought that contact would entail exposure to a variety of grammatical structures and make it possible for the *most* harmonic patterns to diffuse if that were desirable; but here we have two cases where the opposite occurs. Again, is this not a falsification of the principle? Finally, the argument that areality explains non-harmonic patterns would be stronger if a larger survey were done to show that known violations occur only (or at least chiefly) under areal pressure.

Whitman

Does capturing the distinction between the cross-categorial generalizations and the hierarchical and derivational ones require a generative framework with hierarchical phrase structure and movement or mapping between underlying and surface syntax? And do the explanations given for the hierarchical and derivational generalizations require such a framework? More generally, what is the minimum theoretical apparatus required, and how can it be stated framework-neutrally? (If it can be, that is; the demonstration here of de facto categorical absence of OSV languages from data amounting to some 18 percent of the world's languages would seem to be a compelling test case for comparing frameworks and determining what must minimally go into basic theory.)

Hopper

That serialization-like properties arise with the uninflected forms of *take* in this construction is important to the argument. But what was counted—the form *take* or that form only in non-present-tense function? If the positive tokens include non-third-singular presents like *I take his questions and throw them back at him, You always take my questions and treat them as challenges* and not only nonfinite ones like *I try to take his questions and mull them over, You can take this letter and trash it*, then the count may not support the claim as made here, as the former have zero present-tense inflection rather than no inflection. Also, the paper argues that the *take NP and* construction (and with it grammar in general) is emergent. But if serialization-like properties arise precisely in uninflected contexts, isn't there some target or constraint that is guiding their emergence? Finally, how does one get from the claim that serialization in *take NP and* is emergent to the claim that grammar in general is emergent?

In summary, every chapter makes an important and original claim, but except for Garrett's survey of the entirety of Greek and English verb morphology none of them is explicitly enough concerned with replicability and falsifiability to enable linguists

to make one confident that the claims are valid generalizations about language or grammar. This is not a criticism, as hypotheses have to be raised before they can be tested, but I think it is an important next step.

12.3 ARE THERE REALLY ANY UNIVERSALS?

Do universals exist? Or, more precisely, if they exist can we identify them? In addition to the view of hard-wired universals mentioned in section 12.1 above, another view not represented in this book is that, given the small number of language stocks on earth (some two to three hundred) and the need to distribute samples both genealogically and geographically, possible cross-linguistic samples are too small to allow any statistical technique that allows us to view the sample as representing a larger population about which inferences can be drawn (Janssen et al. 2006). Among other things this means that statistical tools do not and cannot enable us to distinguish statistical from categorical universals, and that the statistical standing of universals is no different from that of geographical or temporal clusters. If universals cannot be detected by statistical means and most structural analysis is framework-internal, how are we to know universals?

12.4 FURTHER QUESTIONS

There are three general background questions that I believe need to be made explicit for further work on universals and diachrony:

1. Is cross-categorial harmony more economical than non-harmony? As Albright and Whitman note, non-harmonic systems are not less learnable as first languages. Remarkably consistent harmonic profiles in word and morpheme order appear to have diffused in two cases: head-final order in inner Asia and head-initial order in Mesoamerica. Certainly not every case of areality involves harmonic order (Kuteva and Heine cite examples where areality produces non-harmonic order), but if harmonic order is actually a diffused feature in inner Asia and Mesoamerica and if, as suggested in section 12.2 above, contact involves exposure to a wider range of options and selection from them, then harmonic order may have something to recommend it. Perhaps it eases learning for second-language learners. Or perhaps inner Asia and Mesoamerica are flukes.

2. Do the linguistic types and elements and phenomena attested in the world's known languages exhaust the possible inventory? That is, have we seen everything (or will we have, once the remaining languages are described)? Or are there unknown and unimagined things that are nonetheless possible in language? Evans 1995 shows

that a few hundred years of solitude have been enough to produce some very unusual grammatical properties in Kayardild. Click languages were almost wiped out by the Bantu spread; had they been, we would never have known about clicks and their phonology. These two examples show that there is a good deal of contingency in what has been observed so far, and they suggest that another one or two hundred thousand years of development in pre-Neolithic or at least pre-imperial circumstances might have produced language structures utterly unlike anything known to us now.

3. What is responsible for the structure of utterances in children's speech at the two-word stage? Surely not diachrony, as the grammar of two-word utterances is not transmitted. This example shows that consistencies in structure do occur in the absence of diachrony as a conditioning factor, and it suggests that not all structural universals, or even regularities, are simply the result of language change. That is, I believe the two-word stage calls into question any strong stance to the effect that grammatical structure is necessarily epiphenomenal or emergent and only diachrony has a universal basis.

12.5 CONCLUSION

To summarize what was said in sections 12.2 and 12.3 above, before we can make confident pronouncements on the relationship between language universals and language change, we need to have falsifiable claims and rigorous, replicable, and sizable cross-linguistic surveys of relevant phenomena and their origins. This means that, as in so many other areas of linguistics, progress in this field requires thorough description and documentation of all languages; full comparative-historical description and reconstruction, with explicit accounts of changes, for very many languages and families; and a robust framework-neutral terminology and theoretical apparatus.

This page intentionally left blank

BIBLIOGRAPHY

Adéṣọlá, Olúṣèyẹ Peter (to appear). "Sentence-final *ni*", in Kofi Korankye Saah (ed.), *Niger-Congo Syntax and Semantics 9*. Boston: Boston University African Studies Center.

Ahn, Sang-Cheol (1998). *An Introduction to Korean Phonology*. Seoul: Hanshin Publishing Co.

Aikhenvald, Alexandra Y. (1999). "Serial constructions and verb compounding: Evidence from Tariana (North Arawak)", *Studies in Language* 23: 469–498.

Aissen, Judith (2003). "Differential object marking: Iconicity versus economy", *Natural Language and Linguistic Theory* 21: 435–483.

Albright, Adam (2002a). *The Identification of Bases in Morphological Paradigms*. University of California, Los Angeles, Ph.D. thesis.

—— (2002b). "Islands of Reliability for regular morphology: Evidence from Italian", *Language* 78: 684–709.

—— (2003). "Base selection in analogical change in Yiddish", in Julie Larson and Mary Paster (eds.), *Proceedings of the Twenty-Eighth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, 1–13.

—— (2005). "The morphological basis of paradigm leveling", in Laura J. Downing, T. A. Hall, and Renate Raffelsiefen (eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press, 17–43.

—— (to appear). "Inflectional paradigms have bases too: Arguments from Yiddish", in Asaf Bachrach and Andrew Nevins (eds.), *Inflectional Identity*. Oxford: Oxford University Press.

Albright, Adam and Hayes, Bruce (2002). "Modeling English past tense intuitions with minimal generalization", in Michael Maxwell (ed.), *Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology*. Philadelphia: Association for Computational Linguistics, 58–69.

—— and —— (2003). "Rules vs. analogy in English past tenses: A computational/experimental study", *Cognition* 90: 119–161.

Albright, Adam, Andrade, Argelia Edith, and Hayes, Bruce (2001). "Segmental environments of Spanish diphthongization", in Adam Albright and Taehong Cho (eds.), *UCLA Working Papers in Linguistics, Number 7: Papers in Phonology 5*. Los Angeles: UCLA Department of Linguistics, 117–151.

Alekseev, M. E. (1985). *Voprosy sravnitel'no-istoričeskoj grammatiki lezginskix jazykov: Morfologija, sintaksis*. Moscow: Nauka.

Aleksidze, Zaza, Gippert, Jost, and Schulze, Wolfgang (in preparation). *The Caucasian-Albanian Palimpsest from Mt. Sinai. Edition and Interpretation, with an Introduction by Jean-Pierre Mahé.* [Monumenta Paleographica Medii Aevi.] Turnhout: Brepols.

Allen, Joe and Christiansen, Morten (1996). "Integrating multiple cues in word segmentation: A connectionist model using hints", in Garrison W. Cottrell (ed.), *Proceedings of the Eighteenth Annual Cognitive Science Society Conference*. Mahwah, NJ: Lawrence Erlbaum Associates, 370–375.

Ameka, Felix (1996). "Body parts in Ewe grammar", in Hilary Chappell and William McGregor (eds.), *The Grammar of Inalienability*. Berlin: Mouton de Gruyter, 783–840.

Anderson, Stephen R. (1977). "On mechanisms by which languages become ergative", in Charles N. Li (ed.), *Mechanisms of Syntactic Change*. Austin: University of Texas, 317–363.

—— (1985). *Phonology in the Twentieth Century*. Chicago: University of Chicago.

—— (1989). "Morphological change", in Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge Survey, Volume I: Linguistic Theory: Foundations*. Cambridge: Cambridge University, 324–362.

Andrews, Avery (2001). "Iofu and Spreading Architecture in LFG", in Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG 01 Conference, University of Hong Kong, Hong Kong*. Stanford: CSLI, 1–12.

Anttila, Arto (1997). *Variation in Finnish Phonology and Morphology*. Stanford University, Ph.D. thesis.

Anttila, Arto and Fong, Vivienne (2003). "Variation, ambiguity, and noun classes in English". ROA-589, Rutgers Optimality Archive, http://roa.rutgers.edu/

Archangeli, Diana and Pulleyblank, Douglas (1994). *Grounded Phonology*. Cambridge, Mass.: MIT.

Aristar, Anthony Rodrigues (1991). "On diachronic sources and synchronic patterns: An investigation into the origin of linguistic universals", *Language* 67: 1–33.

Auer, Peter (2000). "On-line Syntax, oder: Was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen", *Sprache und Literatur in Wissenschaft und Unterricht* 31: 43–56.

—— (2005). "Projection in interaction and projection in grammar", *Text* 25: 7–36.

Austin, Peter (1981). "Switch-reference in Australia", *Language* 57.2 (June): 309–334.

—— (1982). "Transitivity and cognate objects in Australian languages", in Paul J. Hopper and Sandra A. Thompson (eds.), *Studies in Transitivity*. New York: Academic, 37–47.

Baayen, R. Harald, Piepenbrock, Richard, and van Rijn, Hedderik (1993). *The CELEX Lexical Database on CD-ROM*. Philadelphia: Linguistic Data Consortium.

Baker, Mark C. (2001). *The Atoms of Language*. New York: Basic Books.

Bakhtin, Mikhail M. (1986). "The problem of speech genres", trans. Vern W. McGee, in Caryl Emerson and Michael Holquist (eds.), *Speech Genres and Other Late Essays*. Austin: University of Texas.

Bamgboṣe, Ayọ (1966). *A Grammar of Yoruba*. Cambridge: Cambridge University.

Barnes, Jonathan (2002). *Positional Neutralization: A Phonologization Approach to Typological Patterns*. University of California, Berkeley, Ph.D. thesis.

—— (2006). *Strength and Weakness at the Interface: Positional Neutralization in Phonetics and Phonology*. Berlin: Mouton de Gruyter.

Baroni, Marco (2000). *Distributional Cues in Morpheme Discovery: A Computational Model and Empirical Evidence*. University of California, Los Angeles, Ph.D. thesis.

Barr, Robin (1994). *A Lexical Model of Morphological Change*. Harvard University, Ph.D. thesis.

Baudouin de Courtenay, Jan (1871/1972). "Some general remarks on linguistics and language", in Edward Stankiewicz (ed. and trans.), *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics* (Indiana University Studies in the History and Theory of Linguistics). Bloomington: Indiana University, 49–80.

Bauman, James J. (1979). "An historical perspective on ergativity in Tibeto-Burman", in Frans Plank (ed.), *Ergativity: Towards a Theory of Grammatical Relations*. London: Academic Press, 419–433.

Benua, Laura (2000). *Phonological Relations Between Words*. New York: Garland.

Berg, René van den (1989). *A Grammar of the Muna Language*. Dordrecht: Foris.

Berman, Howard (1981). "[Review of] *The Languages of Native America: Historical and Comparative Assessment*, ed. by Lyle Campbell and Marianne Mithun", *International Journal of American Linguistics* 47: 248–262.

Bermúdez-Otero, Ricardo and Börjars, Kersti (2006). "Markedness in phonology and syntax: The problem of grounding", *Lingua* 116: 710–756.

Bhatia, Tej (1993). Punjabi: *A Cognitive-Descriptive Grammar*. Routledge.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, and Finegan, Edward (eds.) (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Bickel, Balthasar and Nichols, Johanna (2005). "The semantics of inalienables: A typological survey". Paper presented at the annual meeting of the Linguistic Society of America, 6–9 Jan. 2005, Oakland, California.

Bierwisch, Manfred and Schreuder, Robert (1992). "From concepts to lexical items", *Cognition* 42: 23–60.

Biggs, Bruce (1966). *English–Maori Dictionary*. Wellington: A. H. and A. W. Reed.

Bile, Monique (1988). *Le dialecte crétois ancien*. Paris: Paul Geuthner.

Blake, Barry J. (1977). *Case Marking in Australian Languages* (Linguistics Series 23). Canberra: Australian Institute of Aboriginal Studies.

—— (1987). *Australian Aboriginal Grammar*. London: Croom Helm.

—— (2001). *Case*. Cambridge: Cambridge University.

Blansitt, Edward L. Jr. (1973). "Bitransitive clauses", *Working Papers in Language Universals (Stanford)* 13: 1–26.

Blevins, Juliette (1994). "A phonological and morphological reanalysis of the Maori passive", *Te Reo* 37: 29–53.

—— (1995). "The syllable in phonological theory", in John A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Cambridge, Mass.: Blackwell, 206–244.

—— (1997). "Rules in optimality theory: Two case studies", in I. Roca (ed.), *Derivations and Constraints in Phonology*. Oxford: Clarendon, 27–60.

—— (1999). "Untangling Leti infixation", *Oceanic Linguistics* 38: 383–403.

—— (2001). "Where have all the onsets gone? Initial consonant loss in Australian Aboriginal languages", in Jane Simpson, David Nash, Mary Laughren, and Barry Alpher (eds.), *Forty Years On: Ken Hale and Australian Languages* (Pacific Linguistics 512). Canberra: Australian National University, 481–492.

—— (2003a). "The independent nature of phonotactic constraints: An alternative to syllable-based approaches", in Caroline Féry and Ruben van di Vijver (eds.), *The Syllable in Optimality Theory*. Cambridge: Cambridge University, 375–403.

—— (2003b). "Yurok syllable weight", *International Journal of American Linguistics* 69: 4–24.

—— (2003c). "The phonology of Yurok glottalized sonorants", *International Journal of American Linguistics* 69: 371–396.

—— (2003d). "A note on reduplication in Bugotu and Cheke Holo", *Oceanic Linguistics* 42: 499–505.

Blevins, Juliette, (2004a). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University.

—— (2004b). "The mystery of Austronesian final consonant loss", *Oceanic Linguistics* 43: 179–184.

—— (2004c). "A reconsideration of Yokuts vowels", *International Journal of American Linguistics* 70: 33–51.

—— (2004d). "Klamath sibilant degemination: Implications of a recent sound change", *International Journal of American Linguistics* 70: 279–289.

—— (2005a). "Understanding antigemination: Natural or unnatural history", in Zygmunt Frajzyngier, David Rood, and Adam Hodges (eds.), *Linguistic Diversity and Language Theories*. Amsterdam: Benjamins, 203–234.

—— (2005b). "The role of phonological predictability in sound change: Privileged reduction in Oceanic reduplicated substrings", *Oceanic Linguistics* 44: 455–464.

—— (2006a). "A theoretical synopsis of Evolutionary Phonology", *Theoretical Linguistics* 32: 117–166.

—— (2006b). "Syllable typology", in Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn., vol. 12. Oxford: Elsevier, 333–337.

Blevins, Juliette and Garrett, Andrew (1998). "The origins of consonant-vowel metathesis", *Language* 74: 508–556.

—— and —— (2004). "The evolution of metathesis", in Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.), *Phonetically Based Phonology*. Cambridge: Cambridge University, 117–156.

—— and —— (2007). "The rise and fall of l-sandhi in California Algic", *International Journal of American Linguistics* 73: 72–93.

Bloomfield, Leonard (1933). *Language*. New York: Henry Holt.

Blust, Robert (1978). "Eastern Malayo-Polynesian: A subgrouping argument", in S. A. Wurm and Lois Carrington (eds.), *Proceedings of the Second International Conference on Austronesian Linguistics* (Pacific Linguistics C-61). Canberra: Australian National University, 691–716.

—— (1990). "Three recurrent changes in Oceanic languages", in Jeremy H. C. S. Davidson (ed.), *Pacific Island Languages: Essays in Honour of G. B. Milner*. London: SOAS, 7–28.

—— (1998). "A Lou vocabulary, with phonological notes", in Darrell Tryon (ed.), *Papers in Austronesian Linguistics No. 5* (Pacific Linguistics A-92). Canberra: Australian National University, 35–99.

—— (2000). "Chamorro historical phonology", *Oceanic Linguistics* 39: 83–122.

—— (to appear). *The Austronesian Languages*. Cambridge: Cambridge University.

Boeder, Winfred (1979). "Ergative syntax and morphology in language change: The South Caucasian languages", in Frans Plank (ed.), *Ergativity: Towards a Theory of Grammatical Relations*, 435–480. New York: Academic.

Booij, Geert (to appear). "Lexical storage and phonological change", in Kristin Hanson and Sharon Inkelas (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge, Mass.: MIT.

Bossong, Georg (1985). *Differenzielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr.

—— (1998). "Le marquage différentiel de l'objet dans les langues d'Europe", in Jack Feuillet (ed.), *Actance et valence dans les langues de l'Europe*. Berlin: Mouton de Gruyter, 193–258.

Braune, Wilhelm and Mitzka, Walther (1963). *Althochdeutsche Grammatik* (11th edn.). Tübingen: Niemeyer.

Braunmüller, K. (1991). *Die Skandinavischen Sprachen im Überblick*. Tübingen: Francke Verlag.

Breen, Gavan and Pensalfini, Rob (1999). "Arrernte: A language with no syllable onsets", *Linguistic Inquiry* 30: 1–25.

Brent, Michael (1999). "Speech segmentation and word discovery: A computational perspective", *Trends in Cognitive Science* 3: 294–301.

Brent, Michael and Cartwright, Timothy (1996). "Distributional regularity and phonotactic constraints are useful for segmentation", *Cognition* 61: 93–125.

Bresnan, Joan (2001). *Lexical-Functional Syntax*. Oxford: Blackwell.

Breu, Walter (1994). "Der Faktor Sprachkontakt in einer dynamischen Typologie des Slavischen", in Hans Robert Mehling (ed.), *Slavistische Linguistik 1993*. Munich: Otto Sagner, 41–64.

Broadwell, Aaron (1988). "Reflexive movement in Choctaw", in James P. Blevins and Juli Carter (eds.), *Proceedings of the Eighteenth Meeting of the North East Linguistic Society*. Amherst, Mass.: Graduate Linguistic Student Association, 53–64.

Browman, Catherine P. and Goldstein, Louis M. (1992). "Articulatory phonology: An overview", *Phonetica* 49: 155–180.

Buckley, Eugene (1999). "Uniformity in extended paradigms", in Ben Hermans and Marc van Oostendorp (eds.), *The Derivational Residue in Phonological Optimality Theory*. Amsterdam: Benjamins, 81–104.

Burzio, Luigi (1994). "Metrical consistency", in Eric Ristad (ed.), *Proceedings of the DIMACS Workshop on Human Language*. Providence, RI: American Mathematical Society.

—— (1996). "Surface constraints vs. underlying representation", in Jacques Durand and Bernard Laks (eds.), *Current Trends in Phonology: Models and Methods*. Manchester: European Studies Research Institute, University of Salford, 123–142.

—— (2000). "Segmental contrast meets output-to-output faithfulness", *Linguistic Review* 17: 367–384.

Butt, Miriam (2001). "A reexamination of the accusative to ergative shift in Indo-Aryan", in Miriam Butt and Tracy Holloway King (eds.), *Time Over matter: Diachronic Perspectives on Morphosyntax*. Stanford: CSLI, 105–141.

Bybee, Joan L. (1985). *Morphology: A Study of the Relation Between Meaning and Form*. Amsterdam: Benjamins.

—— (1988a). "Morphology as lexical organization", in Michael Hammond and Michael Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*. San Diego: Academic, 119–141.

—— (1988b). "The diachronic dimension in explanation", in John A. Hawkins (ed.), *Explaining Language Universals*. Oxford: Blackwell, 350–379.

—— (1998). "Usage-based phonology", in Michael Darnell, Edith Moravcsik, Frederick J. Newmeyer, Michael Noonan, and Kathleen M. Wheatley (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*. Amsterdam: Benjamins, 211–242.

—— (2000a). "Lexicalization of sound change and alternating environments", in Michael Broe and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University, 250–268. Repr. in Bybee (2006a), 216–234.

Bybee, Joan L. (2000b). "The phonology of the lexicon: Evidence from lexical diffusion", in Michael Barlow and Suzanne Kemmer (eds.), *Usage-based Models of Language*. Stanford: CSLI, 65–85. Repr. in Bybee (2006a), 199–215.

—— (2001). *Phonology and Language Use*. Cambridge: Cambridge University.

—— (2002). "Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change", *Language Variation and Change* 14: 261–290. Repr. in Bybee (2006a), 235–264.

—— (2003). "Mechanisms of change in grammaticization: The role of frequency", in Richard Janda and Brian D. Joseph (eds.), *Handbook of Historical Linguistics*. Oxford: Blackwell, 602–623.

—— (2006a). *Frequency of Use and the Organization of Language*. Oxford: Oxford University.

—— (2006b). "Language change and universals", in Ricardo Mairal and Juana Gil (eds.), *Linguistic Universals*. Cambridge: Cambridge University, 179–194.

Bybee, Joan and Brewer, Mary (1980). "Explanation in morphophonemics: Changes in Provençal and Spanish preterite forms", *Lingua* 52: 201–242.

Bybee, Joan L. and Moder, Carol Lynn (1983). "Morphological classes as natural categories", *Language* 59: 251–270.

Bybee, Joan L. and Dahl, Östen (1989). "The creation of tense and aspect systems in the languages of the world", *Studies in Language* 13: 51–103.

Bybee, Joan and Scheibman, Joanne (1999). "The effect of usage on degrees of constituency: The reduction of *don't* in English", *Linguistics* 37: 575–596. Repr. in Bybee (2006a), 294–312.

Bybee, Joan L., Perkins, Revere D., and Pagliuca, William (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago.

Cahill, Michael (1999). *Aspects of Morphology and Phonology of Konni*. Ohio State University, Ph.D. thesis.

Cairns, Paul, Shillcock, Richard, Chater, Nick, and Levy, Joe (1997). "Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation", *Cognitive Psychology* 33: 111–153.

Cardinaletti, Anna and Starke, Michael (1999). "The typology of structural deficiency", in Henk van Riemsdijk (ed.), *Clitics in the Languages of Europe*. Berlin: Mouton de Gruyter.

Carstens, Vicki (2002). "Antisymmetry and word order in serial constructions", *Language* 78: 3–50.

Chantraine, Pierre (1961). *Morphologie historique de la langue grecque*. Paris: Klincksieck.

Chao, Yuen-Ren (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California.

Cho, Seung-Bog (1967). *A Phonological Study of Korean: With a Historical Analysis*. Uppsala: Almqvist and Wiksells.

Choe, Hyun-Sook (1988). *Restructuring Parameters and Complex Predicates: A Transformational Approach*. MIT, Ph.D. thesis.

Chomsky, Noam (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

—— (1986). *Barriers*. Cambridge, Mass.: MIT.

Chomsky, Noam and Halle, Morris (1968). *The Sound Pattern of English*. New York: Harper and Row.

Čikobava, Arnold (1938). *Č'anur–megrul–kartuli šedarebiti leksik'oni* [A Laz–Mingrelian–Georgian Comparative Dictionary]. T'pilisi: Mecnierebata Ak'ademiis Sakartvelos Pilialis Gamocema.

Čikobava, Arnold (1942). "Ergat'iuli k'onst'rukciis p'roblemisatvis k'avk'asiur enebši" [The ancient marker of the third-person subject in the Kartvelian languages], *Enimk'is Moambe* 5–6: 13–42.

—— (1943). "P'ermansivis ('xolmeobitis') ist'oriuli adgilisatvis kartuli zmnis uγvlilebis sist'emaši" [The historical position of the permansive (habitual) in the conjugational system of the Georgian verb], *Sbornik materialov dlja opisanija mestnostej i plemen kavkaza* 4: 91–96.

—— (1948). *Ergat'iuli k'onst'rukciis p'roblema iberiul-k'avk'asiur enebši, I* [The problem of the ergative construction in the Ibero-Caucasian languages, I]. Tbilisi: Ak'ademia.

Cinque, Guglielmo (1999). *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University.

—— (2003). "The dual source of adjectives and XP- vs. N raising in the Romance DP". Paper presented at the Incontro annuale di dialettologia, Padua, 26 June 2003; at the Center for Advanced Study in Theoretical Linguistics Conference in Linguistics, Tromsø, 2–4 October 2003; and at the thirty-fourth meeting of the North East Linguistic Society, Stony Brook University, 7–9 November 2003.

—— (2005). "Deriving Greenberg's Universal 20 and its exceptions", *Linguistic Inquiry* 36: 315–332.

Clark, Brady (2004). *A Stochastic Optimality-Theoretic Approach to Clause Structure Variation and Change in Middle English*. Stanford University, Ph.D. thesis.

Cole, Jennifer and Hualde, José Ignacio (1998). "The object of lexical acquisition: A UR-free model", in M. Catherine Gruber, Derrick Higgins, Kenneth S. Olson, and Tamra Wysocki (eds.), *The Proceedings from the Panels of the Chicago Linguistic Society's Thirty-Fourth Meeting*. Chicago: Chicago Linguistic Society, 447–458.

Collins, James T. (1982). "Prothesis in the languages of Central Maluku: An argument from proto-Austronesian grammar", in Armam Halim, Lois Carrington, and S. A. Wurm (eds.), *Papers from the Third International Conference on Austronesian Linguistics, Volume 2: Tracing the Travelers* (Pacific Linguistics C-75). Canberra: Australian National University, 187–200.

Comrie, Bernard (1979). "Morphophonemic exceptions and phonetic distance", *Linguistics* 17: 51–60.

—— (1981). "Aspect and voice: Some reflections on perfect and passive", in Philip J. Tedeschi and Annie Zaenen (eds.), *Tense and Aspect* (Syntax and Semantics 14). New York: Academic, 65–78.

—— (1989). *Language Universals and Linguistic Typology*, 2nd edn. Oxford: Blackwell.

—— (2003). "Reconstruction, typology and reality", in Raymond Hickey (ed.), *Motives for Language Change*, 243–257. Cambridge: Cambridge University.

Connine, Cynthia M. (2004). "It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition", *Psychonomic Bulletin and Review*, 11: 1084–1089.

Corbett, Greville (2000). *Number*. Cambridge: Cambridge University.

Coseriu, Eugenio (1996). " 'Tomo y me voy.' Ein Problem vergleichender europäischer Syntax", *Vox Romanica: Annales Helvetici explorandis linguis Romanicis destinati* 25: 13–55.

Cowan, H. K. J. (1952). "De Austronesisch-Papoea'se taalgrens in de Onderafdeling Hollandia", *Tijdschrift Nieuw Guinea* 13: 133–144; 161–178; 201–207.

Craig, Colette (1977). *The Structure of Jacaltec*. Austin: University of Texas.

Croft, William (1990). *Typology and Universals.* Cambridge: Cambridge University.

Croft, William (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: University of Chicago.

—— (2000). *Explaining Language Change: An Evolutionary Approach*. London: Longman.

—— (2003). *Typology and Universals*, 2nd edn. Cambridge: Cambridge University.

Crowley, Terry (1978). *The Middle Clarence Dialects of Bandjalang*. Canberra: Australian Institute of Aboriginal Studies.

Culler, Jonathan (1988). *Framing the Sign: Criticism and its Institutions*. Tulsa: University of Oklahoma.

Dahl, Östen (2004a). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: Benjamins.

—— (2004b). "Definite articles in Scandinavian: Competing grammaticalization processes in standard and non-standard varieties", in Bernd Kortmann (ed.), *Dialectology Meets Typology: Dialect Grammar from a Cross-linguistic Perspective*. Berlin: Mouton de Gruyter, 147–180.

Dahl, Östen and Koptjevskaja-Tamm, Maria (1998). "Alienability splits and the grammaticalization of possessive constructions", in Timo Haukioja (ed.), *Papers from the Sixteenth Scandinavian Conference of Linguistics*. Turku: University of Turku, 38–49.

Davis, Stuart and Kang, Hyunsook (2006). "English loanwords and the word-final [t] problem in Korean", *Language Research* 42.2: 253–274.

Davitiani, Aleksi, Topuria, Varlam, and Kaldani, Maksime (1957). *Svanuri p'rozauli t'ekst'ebi, II: Balskvemouri k'ilo* [Svan Prose Texts, II: The Lower Bal Dialect]. Tbilisi: Ak'ademia.

Dayley, Jon P. (1985). *Tzutujil Grammar* (University of California Publications in Linguistics 107). Berkeley: University of California.

De Boer, Bart (1999). *Self-Organization in Vowel Systems*. Vrije Universiteit, Brussels, Ph.D. thesis.

—— (2001). *The Origins of Vowel Systems*. Oxford: Oxford University.

Deeters, Gerhard (1927). "Armenisch und Südkaukasisch, II", *Caucasica* 4: 1–64.

—— (1930). *Das khartwelische Verbum*. Leipzig: Markert and Petters.

De Lacy, Paul (2002a). *The Formal Expression of Markedness*. University of Massachusetts, Ph.D. thesis.

—— (2002b). "The interaction of tone and stress in Optimality Theory", *Phonology*, 19: 1–32.

—— (2003). "Maximal words and the Maori passive", in John McCarthy (ed.), *Optimality Theory in Phonology: A Reader*. Oxford: Blackwell, 495–512.

DeLancey, Scott (1981). "An interpretation of split ergativity and related patterns", *Language* 57: 626–57.

Delbrück, Berthold (1880/1974). *Introduction to the Study of Language: A Critical Survey of the History and Methods of Comparative Philology of Indo-European Languages*, trans. E. F. K. Koerner. Amsterdam: Benjamins.

Dench, Alan (1982). "The development of an accusative case marking pattern in the Ngayarda languages of Western Australia", *Australian Journal of Linguistics* 2: 43–59.

Derbyshire, Desmond C. (1977). "Word order universals and the existence of OVS languages", *Linguistic Inquiry* 8: 590–599.

Derbyshire, Desmond C. and Pullum, Geoffrey K. (1981). "Object-initial languages", *International Journal of American Linguistics* 47: 192–214.

—— and —— (eds.)(1986). *Handbook of Amazonian Languages*, Volume 1. Berlin: Mouton de Gruyter.

Diessel, Holger (1999). *Demonstratives: Form, Function, and Grammaticalization* (Typological Studies in Language 42). Amsterdam: Benjamins.

Dirr, Adolph (1928). "Udische Texte", *Caucasica* 5: 60–72.

Dixon, R. M. W. (1972). *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University.

—— (1977). *A Grammar of Yidiɲ*. Cambridge: Cambridge University.

—— (1980). *The Languages of Australia*. Cambridge: Cambridge University.

—— (1979). "Ergativity", *Language* 55: 59–138.

—— (1981). "Wargamay", in R. M. W. Dixon and Barry Blake (eds.), *Handbook of Australian Languages 2*. Amsterdam: Benjamins.

—— (1994). *Ergativity.* Cambridge: Cambridge University.

Donaldson, Tamsin (1980). *Ngiyambaa: The Language of the Wangaaybuwan*. Cambridge: Cambridge University.

Donohue, Cathryn (2004). *Morphology Matters: Case Licensing in Basque*. Stanford University, Ph.D. thesis.

Donohue, Mark (2002). "Tobati", in John Lynch, Malcom Ross, and Terry Crowley (eds.), *The Oceanic Languages*. Richmond, Surrey: Curzon, 186–203.

Dressler, Wolfgang (1977). "Morphologization of phonological processes", in Alphonse Juilland (ed.), *Linguistic Studies Offered to Joseph Greenberg*. Saratoga: Anma Libri II, 313–337.

—— (1985). *Morphonology: The Dynamics of Derivation*. Ann Arbor: Karoma.

Dryer, Matthew S. (1988). "Object-Verb order and Adjective-Noun order: Dispelling a myth", *Lingua* 74: 185–217.

—— (1992). "The Greenbergian word order correlations", *Language* 68: 81–138.

—— (1998). "Why statistical universals are better than absolute universals", in Kora Singer, Randall Eggert, and Gregory Anderson (eds.), *Proceedings of the Thirty-Third Regional Meeting of the Chicago Linguistic Society: Papers from the Panels on Linguistic Ideologies in Contact, Universal Grammar, Parameters and Typology, the Perception of Speech and Other Acoustic Signals*. Chicago: Chicago Linguistic Society, 123–145.

—— (2005a). "Order of adposition and noun phrase", in Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University, 346–349.

—— (2005b). "Order of subject, object, and verb", in Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University, 332–335.

—— (2005c). "Position of interrogative phrases in content questions", in Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University, 378–381.

—— (2005d). "Position of polar question particles", in Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.), *The World Atlas of Language Structures*.Oxford: Oxford University, 374–377.

Du Bois, John (1985). "Competing motivations", in John Haiman (ed.), *Iconicity in Syntax*. Amsterdam: Benjamins, 343–365.

Duhoux, Yves (1992). *Le verbe grec ancien: Élements de morphologie et de syntaxe historiques.* Louvain-la-Neuve: Peeters.

Durie, Mark (1985). *A Grammar of Acehnese on the Basis of a Dialect of North Aceh*. Dordrecht: Foris.

Durie, Mark (1997). "Grammatical structures in verb serialization", in Alex Alsina, Joan Bresnan, and Peter Sells (eds.), *Complex Predicates*. Stanford: CSLI, 289–354.

—— (1999). "The temporal mediation of structure and function", in Michael Darnell, Edith Moravcsik, Frederick Newmeyer, Michael Noonan, and Kathleen Wheatley (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*. Amsterdam: Benjamins, 417–443.

Durie, Mark, Daud, Bukhari, and Hasan, Mawardi (1994). "Acehnese", in Cliff Goddard and Anna Wierzbicka (eds.), *Semantic and Lexical Universals: Theory and Empirical Findings*. Amsterdam: Benjamins, 171–201.

Egerod, Søren (1965). "Verb inflection in Atayal", *Lingua* 15: 251–82.

Einenkel, Eugen (1916). *Geschichte der englischen Sprache, II: Historische Syntax*. Strasburg: Trübner.

Elbert, Samuel H. (1988). *Echo of a Culture: A Grammar of Rennell and Bellona* (Oceanic Linguistics Special Publication 22). Honolulu: University of Hawaii.

Emonds, Joseph (1980). "Word order in generative grammar", *Journal of Linguistic Research* 1: 33–54.

Enfield, N. J. (2002). "Ethnosyntax: Introduction", in N. J. Enfield (ed.), *Ethnosyntax: Explorations in Grammar and Culture*. Oxford: Oxford University, 1–30.

Englebretson, Robert (2003). *Searching for Structure: The Problem of Complementation in Colloquial Indonesian Conversation* (Studies in Discourse and Grammar 13). Amsterdam: Benjamins.

Epps, Patience (2007). "From 'wood' to future tense: Nominal origins of the future construction in Hup". Unpub. ms., University of Texas, Austin.

Ernestus, Mirjam and Baayen, R. Harald (2003). "Predicting the unpredictable: Interpreting neutralized segments in Dutch", *Language* 79: 5–38.

Evans, Nicholas D. (1995). *A Grammar of Kayardild, with Historical-Comparative Notes on Tangkic*. Berlin: Mouton de Gruyter.

—— (2003). "Context, culture, and structuration in the languages of Australia", *Annual Review of Anthropology* 32: 13–40.

Everaert, Martin (2001). "Paradigmatic restrictions on anaphors", in Karine Megerdoomian and Leora Anne Bar-el (eds.), *Proceedings of the Twentieth West Coast Conference on Formal Linguistics*. Somerville, Mass.: Cascadilla, 178–191.

Faltz, Leonard M. (1977). *Reflexivization: A Study in Universal Syntax*. University of California, Berkeley, Ph.D. thesis.

—— (1985). *Reflexivization: A Study in Universal Syntax*. New York: Garland.

Fidelholtz, James (1975). "Word frequency and vowel reduction in English", in Robin E. Grossman, L. James San, and Timothy J. Vance (eds.), *Papers from the Eleventh Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 200–213.

Filip, Hana (2000). "The quantization puzzle", in Carol Tenny and James Pustejovsky (eds.), *Events as Grammatical Objects: The Converging Perspectives of Lexical Semantics and Syntax*. Stanford: CSLI, 39–96.

Fillmore, Charles J. (1988). "The mechanisms of 'construction grammar' ", in Shelley Axmaker, Annie Jaisser, and Helen Singmaster (eds.), *Proceedings of the Fourteenth Annual Meeting of Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, 35–55.

Flemming, Edward (2003). "The relationship between coronal place and vowel backness", *Phonology* 20: 335–373.

Foley, James (1972). "Rule precursors and phonological change by meta-rule", in Robert P. Stockwell and Ronald K. S. Macaulay (eds.), *Linguistic Change and Generative Theory*. Bloomington: Indiana University, 96–100.

Forchheimer, Paul (1953). *The Category of Person in Language* Berlin: de Gruyter.

Ford, Cecilia E. (2001). "At the intersection of turn and sequence: Negation and what comes next", in Margaret Selting and Elizabeth Couper-Kuhlen (eds.), *Studies in Interactional Linguistics*. Amsterdam: Benjamins, 51–79.

Friederici, Angela and Wessels, Jeanine (1993). "Phonotactic knowledge of word boundaries and its use in infant speech perception", *Perception and Psychophysics* 54: 287–295.

Friedman, Victor A. (1979). "Toward a typology of status: Georgian and other non-Slavic languages of the Soviet Union", in Paul R. Clyne, William F. Hanks, and Carol L. Hofbauer (eds.), *The Elements: Papers from the Conference on Non-Slavic Languages of the USSR*. Chicago: Chicago Linguistic Society, 339–350.

Gamq'reliʒe, T. V. and Mač'avariani, G. I. (1965). *Sonant'ta sist'ema da ablaut'i Kartvelur enebši* [The System of Sonants and Ablaut in the Kartvelian Languages]. Tbilisi: Mecniereba.

Gardiner, Alan (1932). *The Theory of Speech and Language.* Oxford: Clarendon.

Garrett, Andrew (1990a). "The origin of NP split ergativity", *Language* 66: 261–296.

—— (1990b). "Applicatives and preposition incorporation", in Katarzyna Dziwirek, Patrick Farrell, and Errapel Mejías-Bikandi (eds.), *Grammatical Relations: A Cross-theoretical Perspective*. Stanford: CSLI, 183–198.

—— (2001). "Reduplication and infixation in Yurok: Morphology, semantics, and diachrony", *International Journal of American Linguistics* 67: 264–312.

Garrett, Andrew and Blevins, Juliette (in press). "Morphophonological analogy", in Sharon Inkelas and Kristin Hanson (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge, Mass.: MIT.

Gast, Volker (2006). *The Grammer of Identity: Intensifiers and Reflexives in Germanic Languages*. London: Routledge.

Gensler, Orin (1994). "On reconstructing the syntagm S-Aux-O-V-Other to Proto–Niger-Congo", in Kevin E. Moore, David A. Peterson, and Comfort Wentum (eds.), *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistic Society, Special Session on Historical Issues in African Linguistics*. Berkeley: Berkeley Linguistics Society, 1–20.

Geraghty, Paul A. (1983). *The History of the Fijian Languages* (Oceanic Linguistics Special Publication 19). Honolulu: University of Hawaii.

Gessner, Suzanne and Hansson, Gunnar (2004). "Anti-homophony effects in Dakelh (Carrier) valence morphology", in M. Ettlinger, N. Fleischer, and M. Park-Doob (eds.), *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, Berkeley: Berkeley Linguistics Society, 93–104.

Gick, Bryan (1999). "A gesture-based account of intrusive consonants in English", *Phonology* 16: 29–54.

Gildersleeve, Basil Lanneau (1900). *Syntax of Classical Greek from Homer to Demosthenes.* New York: American Book Company.

Giorgi, Alessandra and Longobardi, Giuseppe (1991). *The Syntax of Noun Phrases: Configuration, Parameters, and Empty Categories*. Cambridge: Cambridge University.

Givón, Talmy (1975). "Serial verbs and syntactic change: Niger-Congo", in Charles N. Li (ed.), *Word Order and Word Order Change*. Austin: University of Texas, 47–112.

Givón, Talmy (1979). *On Understanding Grammar.* New York: Academic.
—— (1994). *Voice and Inversion*. Amsterdam: John Benjamins.
Goddard, Cliff (1982). "Case systems and case marking in Australian languages: A new interpretation", *Australian Journal of Linguistics* 2: 169–196.
Goldberg, Adele (1995). *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago.
Goldsmith, John (2001). "Unsupervised learning of the morphology of a natural language", *Computational Linguistics* 27: 153–198.
Good, Jeffrey C. (2003). *Strong Linearity: Three Case Studies Towards a Theory of Morphosyntactic Templatic Constructions.*University of California, Berkeley, Ph.D. thesis.
Goodenough, Ward H. and Sugita, Hiroshi (1980). *Trukese–English Dictionary*. Philadelphia: American Philosophical Society.
Greenberg, Joseph H. (1963). "Some universals of grammar with particular reference to the order of meaningful elements", in Joseph H. Greenberg (eds.), *Universals of Grammar*. (Referenced version: 2nd edn., 1966). Cambridge, Mass.: MIT, 73–113.
—— (1966a). *Language Universals, with Special Reference to Feature Hierarchies.* The Hague: Mouton.
—— (1966b). "Synchronic and diachronic universals in phonology", *Language* 42: 508–517.
—— (1969). "Some methods of dynamic comparison in linguistics", in Jan Puhvel (ed.), *Substance and Structure of Language.* Berkeley: University of California, 147–203.
—— (1978a). "Diachrony, synchrony and language universals", in Joseph H. Greenberg, Charles A. Ferguson, and Edith Moravcsik (eds.), *Universals of Human Language, Volume I: Method and Theory.* Stanford: Stanford University, 61–91.
—— (1978b). "Typology and cross-linguistic generalization", in Joseph H. Greenberg, Charles A. Ferguson, and Edith Moravcsik (eds.), *Universals of Human Language, Volume I: Method and Theory.* Stanford: Stanford University, 33–59.
—— (1978c). "How does a language acquire gender markers?", in Joseph H. Greenberg, Charles A. Ferguson, and Edith Moravcsik (eds.), *Universals of Human Language, Volume III: Word Structure*. Stanford: Stanford University, 47–82.
—— (1995). "The diachronic typological approach to language", in Masayoshi Shibatani and Theodora Bynon (eds.), *Approaches to Language Typology*. Oxford: Clarendon, 145–166.
Guion, Susan Guignard (1996). *Velar Palatalization: Coarticulation, Perception, and Sound Change*. University of Texas at Austin, Ph.D. thesis.
—— (1998). "The role of perception in the sound change of velar palatalization", *Phonetica* 55: 18–52.
Haarmann, Harald (1976). *Aspekte der Arealtypologie: Die Problematik der europäischen Sprachbünde* (Tübinger Beiträge zur Linguistik 72). Tübingen: Gunter Narr.
Haiman, John (1979). "Conditionals are topics", *Language* 54: 564–589.
—— (1983). "Iconic and economic motivation", *Language* 59: 781–819.
Hale, Kenneth (1962). "Internal relationships in Arandic of Central Australia", in Arthur Capell (ed.), *Some Linguistic Types in Australia* (Oceania Linguistic Monograph 7). Sydney: University of Sydney, 171–183.
—— (1964). "Classification of Northern Paman languages, Cape York Peninsula, Australia: A research report", *Oceanic Linguistics* 3: 248–265.

Hale, Kenneth (1973a). "Deep-surface canonical disparities in relation to analysis and change: An Australian example", in Thomas A. Sebeok (ed.), *Current Trends in Linguistics 11*. The Hague: Mouton, 401–458.
—— (1973b). "A note on subject–object inversion in Navajo", in Braj B. Kachru, Robert B. Lees, Yakov Malkiel, Angelina Pietrangeli, and Sol Saporta (eds.), *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane*. Urbana: University of Illinois, 300–309.
—— (1976). "Phonological developments in a Northern Paman language: Uradhi", in Peter Sutton (ed.), *Languages of Cape York* (Australian Aboriginal Studies Research and Regional Studies 6). Canberra: Australian Institute of Aboriginal Studies, 41–49.
Hall, Tracy Alan (1989). "Lexical phonology and the distribution of German [ç] and [x]", *Phonology* 6: 1–17.
Hansson, Gunnar (2004). "Long-distance voicing agreement: An evolutionary perspective", in Marc Ettlinger, Nicholas Fleisher, and Mischa Park-Doob (eds.), *Proceedings of the Thirtieth Annual Meeting of the Berkeley Linguistics Society.* Berkeley: Berkeley Linguistics Society, 130–141.
Harris, Alice C. (1981). *Georgian Syntax: A Study in Relational Grammar.* Cambridge: Cambridge University.
—— (1985). *Diachronic Syntax: The Kartvelian Case* (Syntax and Semantics 18). New York: Academic.
—— (2002a). *Endoclitics and the Origins of Udi Morphosyntax*. Oxford: Oxford University.
—— (2002b). "On the origins of circumfixes in Kartvelian", in Wolfram Bublitz, Manfred von Roncador, and Heinz Vater (eds.), *Philologie, Typologie und Sprachstruktur: Festschrift für Winfried Boeder zum 65. Geburtstag*. Frankfurt am Main: Peter Lang, 305–322.
—— (2005). "The challenge of typologically unusual structures", in Geert Booij, Emiliano Guevara, Angela Ralli, Salvatore Sgroi, and Sergio Scalise (eds.), *Morphology and Linguistic Typology*, On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4), Catania, 21–23 September 2003, 277–284. http://morbo.lingue.unibo.it/mmm/ (ISSN number 1826–7491).
—— (in press). "Origins of differential unaccusative/unergative case marking: Implications for innateness", in Donna Gerdts, John Moore, and Maria Polinsky (eds.), *Festschrift for David Perlmutter*. Cambridge, Mass.: MIT.
Harris, Alice C. and Campbell, Lyle (1995). *Historical Syntax in Cross-Linguistic Perspective.* Cambridge: Cambridge University.
—— and —— (1996). "Syntactic doublets and language change". Unpub. ms. presented at the 1996 Annual Meeting of the Linguistic Society of America, San Diego.
Harris, Alice C. and Xu, Zheng (2006). "Diachronic morphological typology", in Keith Brown (ed.), *Encyclopedia of Language and Linguistics*. Oxford: Elsevier, 509–515.
Haspelmath, Martin (1993a). *A Grammar of Lezgian* (Mouton Grammar Library 9). Berlin: Mouton de Gruyter.
—— (1993b). "More on the typology of inchoative/causative verb alternations", in Bernard Comrie and Maria Polinsky (eds.), *Causatives and Transitivity* (Studies in Language Companion Series 23). Amsterdam: Benjamins, 87–120.
—— (1994). "Passive participles across languages", in Barbara Fox and Paul J. Hopper (eds.), *Voice: Form and Function* (Typological Studies in Language 27). Amsterdam: Benjamins, 151–177.

Haspelmath, Martin (1999a). "Optimality and diachronic adaptation", *Zeitschrift für Sprachwissenschaft* 18: 180–205.
—— (1999b). "Explaining article-possessor complementarity: Economic motivation in noun phrase syntax", *Language* 75: 227–243.
—— (1999c). "On the cross-linguistic distribution of same-subject and different-subject complement clauses: Economic vs. iconic motivation'. Paper presented at the International Conference on Cognitive Linguistics, Stockholm, July 1999. http://email.eva.mpg.de/~haspelmt/papers.html
—— (2002). *Understanding Morphology.* London: Arnold.
—— (2004a). "Explaining the ditransitive person-role constraint: A usage-based approach", *Constructions* 2/2004. urn:nbn:de:0009-4-359 (http://www.constructions-online.de)
—— (2004b). "Does linguistic explanation presuppose linguistic description?", *Studies in Language* 28: 554–579.
—— (2004c). "Coordinating constructions: An overview", in Martin Haspelmath (ed.), *Coordinating Constructions* (Typological Studies in Language 58). Amsterdam: Benjamins.
—— (2006). "Against markedness (and what to replace it with)", *Journal of Linguistics* 42: 25–70.
—— (2007a). "A frequentist explanation of some universals of reflexive marking". Unpub. ms., Max Planck Institute for Evolutionary Anthropology, Leipzig. http://email.eva.mpg.de/~haspelmt/papers.html
—— (2007b). "Frequency vs. iconicity in explaining grammatical asymmetries". Unpub. ms., Max Planck Institute for Evolutionary Anthropology, Leipzig. http://email.eva.mpg.de/~haspelmt/papers.html
—— (2008). "Parametric versus functional explanations of syntactic universals". Unpub. ms., Max Planck Institute for Evolutionary Anthropology, Leipzig; to appear in a volume edited by Theresa Biberauer and Anders Holmberg. http://email.eva.mpg.de/~haspelmt/publist.html
—— Dryer, Matthew S., Gil, David, and Comrie, Bernard (eds.) (2005). *The World Atlas of Language Structures*. Oxford: Oxford University.
Hawkins, John A. (1983). *Word Order Universals*. New York: Academic.
—— (ed.) (1988a). *Explaining Language Universals*. Oxford: Basil Blackwell.
—— (1988b). "Introduction", in John A. Hawkins (ed.), *Explaining Language Universals*. Oxford: Basil Blackwell, 3–28.
—— (1999). "Processing complexity and filler-gap dependencies across grammars", *Language* 75.2: 244–285.
—— (2004). *Efficiency and Complexity in Grammars.* Oxford: Oxford University.
Hayes, Bruce (1995). "On what to teach the undergraduates: Some changing orthodoxies in phonological theory", in Ik-Hwan Lee (ed.), *Linguistics in the Morning Calm 3*. Seoul: Hanshin, 59–77.
—— (1999). "Phonological restructuring in Yidiɲ and its theoretical consequences", in Ben Hermans and Marc van Oostendorp (eds.), *The Derivational Residue in Phonology*. Amsterdam: Benjamins, 175–205.
—— (2004). "Phonological acquisition in optimality theory: The early stages", in René Kager, Joe Pater, and Wim Zonneveld (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: Cambridge University.

Hayes, Bruce and Steriade, Donca (2004). "Introduction: The phonetic bases of phonological markedness", in Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.), *Phonetically Based Phonology*. Cambridge: Cambridge University, 1–33.

Heine, Bernd (2003). "Grammaticalization", in Brian D. Joseph and Richard D. Janda (eds.), *The Handbook of Historical Linguistics.* Oxford: Blackwell, 575–601.

Heine, Bernd and Kuteva, Tania (2001). "On context and concretization in the rise of German clause connectives", *Sprachtypologie und Universalienforschung* 2: 451–467.

—— and —— (2002). *World Lexicon of Grammaticalization.* Cambridge: Cambridge University.

—— and —— (2003). "Contact-induced grammaticalization", *Studies in Language* 27: 529–572.

—— and —— (2005). *Language Contact and Grammatical Change.* Cambridge: Cambridge University.

—— and —— (2006). *The Changing Languages of Europe.* Oxford: Oxford University.

—— and —— (2007). "On personal pronouns". Unpub. ms., University of Cologne.

Heine, Bernd and Reh, Mechthild (1984). *Grammaticalization and Reanalysis in African Languages*. Hamburg: Helmut Buske Verlag.

Heine, Bernd, Claudi, Ulrike, and Hünnemeyer, Friederike (1991). *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago.

Held, Warren H., Jr. (1957). *The Hittite Relative Sentence* (Language Dissertation 55). Baltimore: Linguistic Society of America.

Herring, Susan (1991). "The grammaticalization of rhetorical questions in Tamil", in Elizabeth Traugott and Bernd Heine (eds.), *Approaches to Grammaticalization, Volume I*. Amsterdam: Benjamins, 253–285.

Hestvik, Arild (1992). "LF-movement of pronouns and antisubject orientation", *Linguistic Inquiry* 23: 557–94.

Hickey, Raymond (ed.) (2003). *Motives for Language Change*. Cambridge: Cambridge University.

Hino, Yasushi and Lupker, Stephen J. (1996). "Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts", *Journal of Experimental Psychology: Human Perception and Performance* 22: 1331–1356.

Hoberman, Robert D. (1988). "The history of the Modern Aramaic pronouns and pronominal suffixes", *Journal of the American Oriental Society* 108: 557–575.

Hock, Hans Henrich (1991). *Principles of Historical Linguistics*, 2nd edn. Berlin: Mouton de Gruyter.

Hock, Hans Henrich and Joseph, Brian D. (1996). *Language History, Language Change, and Language Relationship: An Introduction to Historical Comparative Linguistics.* Berlin: Mouton de Gruyter.

Hockett, Charles F. (1942). "A system of descriptive phonology", *Language* 18: 3–21.

Hohepa, Patrick W. (1967). *A Profile Generative Grammar of Maori* (Indiana University Publications in Anthropology and Linguistics, Memoir 20). Baltimore: Waverly.

Holisky, Dee Ann (1981). *Aspect and Georgian Medial Verbs*. Delmar, NY: Caravan.

Holmberg, Anders (1986). *Word Order and Syntactic Features in the Scandinavian Languages and English*. University of Stockholm, Ph.D. thesis.

Hook, Peter E. (1974). *The Compound Verb in Hindi.* Ann Arbor: University of Michigan Center for Southeast Asian Studies.

Hooper, Joan B. (1976a). *Introduction to Natural Generative Phonology*. New York: Academic Press.

Hooper, Joan B. (1976b). "Word frequency in lexical diffusion and the source of morphophonological change", in William Christie (ed.), *Current Progress in Historical Linguistics*. Amsterdam: North Holland, 96–105. Repr. in Bybee (2006a), 23–34.

Hopper, Paul J. (1986). "Causes and affects", in William H. Eilfort, Paul Kroeber, and Karen L. Peterson (eds.), *Papers from the Parasession on Causatives and Agentivity at the Twenty-First Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 67–88.

—— (1987). "Emergent grammar", in Jon Aske, Natasha Beery, Laura A. Michaelis, and Hana Filip (eds.), *The Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, 139–157.

Hopper, Paul J. (1992). "Times of the sign: On temporality in recent linguistics", *Time and Society* 1: 223–238.

—— (1998). "Emergent grammar", in Michael Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Linguistic Structure*. Hillsdale, NJ: Lawrence Erlbaum.

—— (2001). "Grammatical constructions and their discourse origins: Prototype or family resemblance?", in Martin Pütz and Susanna Niemeier (eds.), *Applied Cognitive Linguistics: Theory, Acquisition, and Language Pedagogy*. Berlin: Mouton de Gruyter, 109–130.

—— (2002). "Hendiadys and auxiliation in English", in Joan Bybee and Michael Noonan (eds.), *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson*. Amsterdam: Benjamins, 145–173.

—— (to appear). "The openness of grammatical constructions", *Proceedings of the Fortieth Annual Meeting of the Chicago Linguistic Society.*

Hopper, Paul J. and Thompson, Sandra (1980). "Transitivity in grammar and discourse", *Language* 56: 251–299.

Hopper, Paul J. and Traugott, Elizabeth Closs (1993). *Grammaticalization*. Cambridge: Cambridge University.

Huang, Yan (2000). *Anaphora: A Cross-Linguistic Study.* Oxford: Oxford University.

Hyman, Larry M. (1975). *Phonology: Theory and Analysis.* New York: Holt, Rinehart, and Winston.

—— (1977). "Phonologization", in Alphonse G. Juilland (ed.), *Linguistic Studies Offered to Joseph Greenberg on the Occasion of His Sixtieth Birthday*. Saratoga, Calif.: Anma Libri, 407–418.

Imedadze, Natela and Tuite, Kevin (1992). "The acquisition of Georgian", in Dan Isaac Slobin (ed.), *The Crosslinguistic Study of Language Acquisition*, Volume 3. Hillsdale, NJ: Lawrence Erlbaum, 39–109.

Ingram, David (1976). *Phonological Disability in Children*. London: Edward Arnold.

Iverson, Gregory K. and Salmons, Joseph C. (2005). "Filling the gap: English tense vowel plus final /ʃ/", *Journal of English Linguistics* 33: 207–221.

Jacobson, Steven A. (1984). *Yup'ik Eskimo Dictionary*. Fairbanks: Alaska Native Language Center.

Jäger, Gerhard (2004). "Learning constraint sub-hierarchies: The Bidirectional Gradual Learning Algorithm", in Reinhard Blutner and Henk Zeevat (eds.), *Optimality Theory and Pragmatics.* Basingstoke, Hants: Palgrave Macmillan, 251–287.

Jäger, Gerhard and Rosenbach, Annette (2003). "Evolutionary OT and the emergence of possession splits". Paper presented at the workshop on Logic, Neural Networks, and OT, Zentrum für Allgemeine Sprachwissenschaft, Berlin, July 2003.

Jakobson, Roman (1929/1962). "Remarqes sur l'évolution phonologique du russe comparée à celle des autres langues slaves'. *Travaux du Cercle Linguistique de Prague 2*. Repr. in

Roman Jakobson (1962). *Selected Writings I: Phonological Studies*. The Hague: Mouton, 7–116.

—— (1932). "Zur Struktur des russischen Verbums", in *Charisteria Guilelmo Mathesio quinquagenari a discipulis et Circuli Linguistici Pragensis sodalibus oblata*. Prague: Sumptibus 'Prazsky Linguisticky Krouzek", 74–84.

—— (1939/1971). "Signe zéro", in *Mélanges de linguistique offerts à Charles Bally*, 143–152. Repr. in Roman Jakobson (1971). *Selected Writings II: Word and Language*. The Hague: Mouton, 211–219.

Janda, Richard (1999). "Accounts of phonemic split have been greatly exaggerated—But not enough", in *Proceedings of the Fourteenth International Congress of Phonetic Sciences*, San Francisco, 329–332.

Janssen, Dirk P., Bickel, Balthasar, and Zúñiga, Fernando (2006). "Randomization tests in language typology", *Linguistic Typology* 10: 419–440.

Jeffers, Robert J. and Zwicky, Arnold M. (1980). "The evolution of clitics", in Elizabeth Closs Traugott, Rebecca Labrum, and Susan Shepherd (eds.), *Papers from the Fourth International Conference on Historical Linguistics* Amsterdam: Benjamins, 221–232.

Jespersen, Otto (1924). *The Philosophy of Grammar.* New York: Norton.

—— (1933). *Essentials of English Grammar.*. New York: Holt.

—— (1942). *A Modern English Grammar on Historical Principles, Part VI: Morphology*. London: George Allen and Unwin.

Johnson, Keith (1997). "Speech perception without speaker normalization", in Keith Johnson and John W. Mullennix (eds.), *Talker Variability in Speech Processing.* San Diego: Academic, 145–165.

Joos, Martin (1962). *The Five Clocks*. Bloomington, Ind.: Publications of the Indiana University Research Center in Anthropology, Folklore, and Linguistics 22. *International Journal of American Linguistics* 28:2, part 5.

Julien, Marit (2003). "Word order type and syntactic structure", in Johan Rooryck and Pierre Pica (eds.), *Linguistic Variation Yearbook, Volume 1, 2001*. Amsterdam: John Benjamins, 17–59.

Jusczyk, Peter, Friederici, Angela, Wessels, Jeanine, Svenkerud, Vigdis, and Jusczyk, Ann Marie (1993). "Infants' sensitivity to the sound patterns of native language words", *Journal of Memory and Language* 32: 402–420.

Jusczyk, Peter, Luce, Paul, and Charles-Luce, Jan (1994). "Infants' sensitivity to phonotactic patterns in the native language", *Journal of Memory and Language* 33: 630–645.

Kader, Mashudi (1976). *The Syntax of Malay Interrogatives*. Simon Fraser University, Burnaby, BC, Ph.D. thesis.

Kager, Rene (1999). *Optimality Theory.* Cambridge: Cambridge University.

—— (in press). "Lexical irregularity and the typology of contrast", In Kristin Hanson and Sharon Inkelas (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge, Mass.: MIT.

Kang, Yoonjung (2003). "Sound changes affecting noun-final coronal obstruents in Korean", in W. McClure (ed.), *Japanese/Korean Linguistics, Volume 12*. Stanford: CSLI.

Katada, Fusa (1991). "The LF representation of anaphors", *Linguistic Inquiry* 22: 287–313.

Kavitskaya, Darya (2002). *Compensatory Lengthening: Phonetics, Phonology, Diachrony.* New York: Routledge.

Kavtaraʒe, Ivane (1954). *Zmnis ʒiritadi k'at'egoriebis ist'oriisatvis ʒvel kartulši* [On the History of Basic Verbal Categories in Old Georgian]. Tbilisi: Ak'ademia.
Kayne, Richard (1994). *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT.
—— (1997). "The English complementizer *of*", *Journal of Comparative Germanic Linguistics* 1: 43–54.
Kazenin, Konstantin I. (1994). "Focus constructions in Daghestanian languages and the typology of focus constructions". Unpub. ms., Moscow State University.
—— (1995). "Focus constructions in North Caucasian languages", Unpub. ms., Moscow State University.
—— (2002). "Focus in Daghestanian and word order typology", *Linguistic Typology* 6: 289–316.
Keenan, Edward L. and Comrie, Bernard (1977). "Noun phrase accessibility and universal grammar", *Linguistic Inquiry* 8: 63–99.
Keller, Rudi (1990/1994). *Sprachwandel: Von der unsichtbaren Hand in der Sprache*. Tübingen: Francke.
—— (1994). *On Language Change: The Invisible Hand in Language*. London: Routledge. Eng. trans. of Keller (1990/1994).
Kemmer, Suzanne (1993). *The Middle Voice.* Amsterdam: Benjamins.
Kenstowicz, Michael (1996). "Base identity and uniform exponence: Alternatives to cyclicity", in Jacques Durand and Bernard Laks (eds.), *Current Trends in Phonology: Models and Methods*. Manchester: European Studies Research Institute, University of Salford, 363–394.
Kenstowicz, Michael and Kisseberth, Charles (1977). *Topics in Phonological Theory*. New York: Academic.
Kibre, Nicholas (1998). "Formal property inheritance and consonant/zero alternations in Maori verbs". ROA-285, Rutgers Optimality Archive, http://roa.rutgers.edu/.
Kim, Heung-Gyu and Kang, Beom-Mo (2000). "Frequency analysis of Korean morpheme and word usage". Technical report, Seoul: Institute of Korean Culture, Korea University.
Kim, Hyunsoon (1999). "The place of articulation of Korean affricates revisited", *Journal of East Asian Linguistics* 8: 313–347.
—— (2001). "A phonetically based account of phonological stop assibilation", *Phonology* 18: 81–108.
Kim-Renaud, Young-Key (1974). *Korean Consonontal Phonology*. University of Hawaii, Ph.D. thesis.
King, Robert (1969). *Historical Linguistics and Generative Grammar.* Englewood Cliffs, NJ: Prentice-Hall.
—— (1980). "The history of final devoicing in Yiddish", in Marvin I. Herzog, Barbara Kirshenblatt-Gimblett, Dan Miron, and Ruth Wisse (eds.), *The Field of Yiddish: Studies in Language, Folklore, and Literature*. Philadelphia: Institute for the Study of Human Issues, 371–430.
Kiparsky, Paul (1968). "Linguistic universals and linguistic change", in Emmon Bach and Robert T. Harms (eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, 171–202.
—— (1972). "Explanation in phonology", in Stanley Peters (ed.), *Goals in Linguistic Theory*. Englewood Cliffs, NJ: Prentice-Hall. Repr. in Kiparsky (1982), ch. 5.
—— (1978). "Analogical change as a problem for linguistic theory", in Braj B. Kachru (ed.), *Linguistics in the Seventies: Directions and Prospects*, *Studies in the Linguistic Sciences* 8: 72–96.
—— (1982). *Explanation in Phonology*. Dordrecht: Foris.

Kiparsky, Paul (1985). “Some consequences of lexical phonology”, *Phonology Yearbook* 2: 85–138.
—— (1988). “Phonological change”, in Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge Survey, Volume I, Linguistic Theory: Foundations*. Cambridge: Cambridge University, 363–415.
—— (1992). “Analogy”, in William Bright (ed.), *International Encyclopedia of Linguistics, Volume 1*. New York: Oxford University, 56–61.
—— (1995). “The phonological basis of sound change”, in John A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Cambridge, Mass.: Blackwell, 640–670.
—— (1997). “The rise of positional licensing”, in Ans van Kemenade and Nigel Vincent (eds.), *Parameters of Morphosyntactic Change*. Oxford: Oxford University, 460–494.
—— (2001). “Structural case in Finnish”, *Lingua* 111: 315–376.
—— (2002). “Disjoint reference and the typology of pronouns”, in Ingrid Kaufmann and Barbara Stiebels (eds.), *More than Words: A Festschrift for Dieter Wunderlich*. Berlin: Akademie Verlag, 179–226.
—— (2003). “Finnish noun inflection”, in Diane Nelson and Satu Manninen (eds.), *Generative Approaches to Finnic and Saami Linguistics*. Stanford: CSLI, 109–161.
Kirby, Simon (1999). *Function, Selection, and Innateness: The Emergence of Language Universals.* Oxford: Oxford University.
Klenin, Emily (1983). *Animacy in Russian: A New Interpretation* (UCLA Slavic Studies 6). Columbus, Ohio: Slavica.
Klokeid, Terry J. (1978). “Nominal inflection in Pamanyungan: A case study in relational grammar”, in Werner Abraham (ed.), *Valence, Semantic Case, and Grammatical Relations*. Amsterdam: Benjamins, 577–615.
Ko, Heejeong (2006). “Base-output correspondence in Korean nominal inflection”, *Journal of East Asian Linguistics* 15: 195–243.
Ko, Kwang-Mo (1989). “Explaining the noun-final change of t > s”, *Eoneohak* 11: 3–22.
Koch, Harold (1997). “Pama-Nyungan reflexes in Arandic languages”, in Darrell Tryon and Michael Walsh (eds.), *Boundary Rider: Essays in Honour of Geoffrey O’Grady* (Pacific Linguistics C-136). Canberra: Australian National University, 271–302.
König, Ekkehard and Haspelmath, Martin (1999). “Der europäische Sprachbund”, in Norbert Reiter (ed.), *Eurolinguistik: Ein Schritt in die Zukunft*. Wiesbaden: Harrassowitz, 111–127.
König, Ekkehard and Siemund, Peter (1999). “Intensifiers and reflexives: A typological perspective”, in Zygmunt Frajzyngier and Traci S. Curl (eds.), *Reflexives: Forms and Functions* (Typological Studies in Language 40). Amsterdam: Benjamins, 41–74.
Koopman, Hilda (2005). “Korean (and Japanese) morphology from a syntactic perspective”, *Linguistic Inquiry* 36: 601–633.
Koptjevskaja-Tamm, Maria (1996). “Possessive noun phrases in Maltese: Alienability, iconicity, and grammaticalization”, *Rivista di Linguistica* 8: 245–274.
—— (1993). *Nominalizations*. London: Routledge.
Kortmann, Bernd (ed.) (2004). *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective* (Trends in Linguistics 153). Berlin: Mouton de Gruyter.
Krejnovič, E. A. (1958). *Jukagirskij Jazyk*. Moscow and Leningrad: Akademija Nauk.
Kroch, Anthony (1989a). “Function and grammar in the history of English: Periphrastic *do*”, in Ralph W. Fasold and Deborah Schiffrin (eds.), *Language Change and Variation*. Amsterdam: Benjamins, 133–172.

Kroch, Anthony (1989b). "Reflexes of grammar in patterns of language change", *Language Variation and Change* 1: 199–244.

Kroch, Anthony and Taylor, Ann (2000). "Verb-object order in Early Middle English", in Susan Pintzuk, George Tsoulas, and Anthony Warner (eds.), *Diachronic Syntax: Models and Mechanisms*. Oxford: Oxford University, 132–163.

Krygier, Marcin (1997). *From Regularity to Anomaly: Inflectional i-umlaut in Middle English* (Bamberger Beiträge zur Englischen Sprachwissenschaft 40). Frankfurt am Main: Peter Lang.

Kuryłowicz, Jerzy (1945–49/1995). "La nature des procès dits analogiques", *Acta Linguistica* 5: 15–37. Eng. trans. with intro. by Margaret Winters (1995), *Diachronica* 12: 113–145.

—— (1964). *The Inflectional Categories of Indo-European*. Heidelberg: Carl Winter.

Kuteva, Tania (1998). "On identifying an evasive gram: Action narrowly averted", *Studies in Language* 22.1: 113–160.

—— (2001). *Auxiliation: An Enquiry into the Nature of Grammaticalization*. Oxford: Oxford University.

—— (forthcoming). "On the 'frills' of grammaticalization", in Elena Seoane (ed.), *New Reflections on Grammaticalization, Volume 3*. Amsterdam: Benjamins.

Kuteva, Tania and Heine, Bernd (forthcoming). *An Integrative Model of Grammaticalization*.

Labov, William (2001). *Principles of Linguistic Change, Volume II: Social Factors.* Oxford: Blackwell.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites.* Stanford: Stanford University.

Lass, Roger (1990). "How to do things with junk: Exaptation in language evolution", *Journal of Linguistics* 26: 79–102.

Laury, Ritva (1997). *Demonstratives in Interaction: The Emergence of a Definite Article in Finnish* (Studies in Discourse and Grammar 7). Amsterdam: Benjamins.

Lazard, Gilbert (1986). "Les prépositions *pa(d)* et *bē (ō)* en persan et en pehlevi", in Rüdiger Schmitt and Prods Oktor Skjaervø (eds.), *Studia Grammatica Iranica: Festschrift für Helmut Humbach*. München: R. Kitzinger, 245–255.

—— (2001). "Le marquage différentiel de l'objet", in Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), *Language Typology and Language Universals: An International Handbook*, Volume 2. Berlin: Walter de Gruyter, 873–885.

Lee, Hyuck-Joon (n.d.). "An analysis of Korean stem final consonants'. Unpub. ms., University of California, Los Angeles.

Lee, Insook (1999). *A Principles-and-Parameters Approach to the Acquisition of (the Morphosyntax of) IP in Korean*. University of Essex, Ph.D. thesis.

Lee, June-Yub (1994). "Hcode: Hangul Code Conversion Program, Version 2.1'. ftp://ftp.kreonet.re.kr/pub/hangul/cair-archive/code/hcode/.

Leech, Geoffrey, Rayson, Paul, and Wilson, Andrew (2001). *Word Frequencies in Written and Spoken English based on the British National Corpus.* Harlow: Pearson Education.

Lefebvre, Claire and Brousseau, Anne-Marie (2002). *A Grammar of Fongbe.* Berlin: Mouton de Gruyter.

Lehmann, Christian (1982/1995). *Thoughts on Grammaticalization: A Programmatic Sketch* (Arbeiten des Kolner Universalien-Projekts 48). Köln: Institut für Sprachwissenchaft. Repr. (1995) Munich: LINCOM.

Leopold, Werner F. (1948). "German *ch*", *Language* 24: 179–180.

Lichtenberk, Frantisek (1991). "Semantic change and heterosemy in grammaticalization", *Language* 67: 475–509.

—— (2001). "On the morphological status of thematic consonants in two Oceanic languages", in Joel Bradshaw and Kenneth L. Rehg (eds.), *Issues in Austronesian Morphology: A Focusschrift for Byron W. Bender* (Pacific Linguistics 519). Canberra: Australian National University, 123–147.

Lightfoot, David W. (1979). *Principles of Diachronic Syntax*. Cambridge: Cambridge University.

—— (1991). *How to Set Parameters: Arguments from Language Change*. Cambridge, Mass.: MIT.

—— (1999). *The Development of Language.* Cambridge, Mass: MIT.

—— (ed.) (2002). *Syntactic Effects of Morphological Change*. Oxford: Oxford University.

—— (2003). "Grammaticalisation: Cause or effect?", In Raymond Hickey (ed.), *Motives for Language Change*. Cambridge: Cambridge University, 99–122.

Lindblom, Björn (1992). "Phonological units as adaptive emergents of lexical development", in Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*. Timonium, MD: York, 131–163.

Lindblom, Björn, MacNeilage, Peter and Studdert-Kennedy, Michael (1984). "Self-organizing processes and the explanation of language universals", in Brian Butterworth, Bernard Comrie, and Östen Dahl (eds.), *Explanations for Language Universals*. Berlin: Walter de Gruyter, 181–203.

Linell, Per (2005). *The Written Language Bias in Linguistics: Its Nature, Origins, and Transformations.* London: Routledge.

Longobardi, Giuseppe (1994). "Reference and proper names: A theory of N-movement in syntax and logical form", *Linguistic Inquiry* 25: 609–665.

—— (2001). "The structure of DPs: Some principles, parameters, and problems", in Mark Baltin and Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell, 562–603.

Lord, Albert (1960). *The Singer of Tales.* Cambridge, Mass.: Harvard University.

Lord, Carol (1982). "The development of object markers in serial verb languages", In Paul J. Hopper and Sandra A. Thompson (eds.), *Studies in Transitivity*. Amsterdam: Benjamins, 277–300.

—— (1993). *Historical Change in Serial Verb Constructions* (Typological Studies in Language 26). Amsterdam: Benjamins.

Lord, Carol and Craig, Louisa Benson (2004). "Conjunction and concatenation in Sgaw Karen: Familiarity, frequency, and conceptual unity", in Martin Haspelmath (ed.), *Coordinating Constructions* (Typological Studies in Language 58). Amsterdam: Benjamins, 357–371.

Luick, Karl (1914–40). *Historische Grammatik der englischen Sprache,* 2 vols. Stuttgart: B. Tauchnitz.

Lynch, John (2000). *A Grammar of Anejom* (Pacific Linguistics 507). Canberra: Australian National University.

MacDonald, Lorna (1990). *A Grammar of Tauya*. Berlin: Mouton de Gruyter.

MacFarland, Talke and Pierrehumbert, Janet (1991). "On ich-Laut, ach-Laut and Structure Preservation", *Phonology* 8: 171–180.

McCarthy, John J. (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University.

—— (2003). "Comparative markedness", *Theoretical Linguistics* 29: 1–51.

—— (2005). "Optimal paradigms", in Laura Downing, Tracy Alan Hall, and Renate Raffelsiefen (eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University, 170–210.

McCarthy, John and Prince, Alan (1994). "The emergence of the unmarked: Optimality in prosodic morphology", in Merce Gonzalez (ed.), *Proceedings of the Twenty-Fourth Meeting of the North East Linguistic Society*. Amherst, Mass.: Graduate Linguistic Student Association, 333–379.

—— and —— (1995). "Faithfulness and reduplicative identity", in Jill N. Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory* (University of Massachusetts Occasional Papers in Linguistics 18). Amherst, Mass.: Graduate Linguistic Student Association, 249–384.

McCloskey, James (1991). "Clause structure, ellipsis, and proper government in Irish", *Lingua* 85: 259–302.

McConvell, Patrick (1981). "How Lardil became accusative", *Lingua* 55: 141–179.

McDaniel, Dana (1989). "Partial and multiple wh-movement", *Natural Language and Linguistic Theory* 7: 565–604.

McGregor, William (1996). "The grammar of nominal prefixing in Nyulnyul", in Hilary Chappell and William McGregor (eds.), *The Grammar of Inalienability.* Berlin: Mouton de Gruyter, 251–292.

McMahon, April (2000). *Change, Chance, and Optimality.* Oxford: Oxford University.

McWhorter, John H. (1998). "Identifying the creole prototype: Vindicating a typological class", *Language* 74: 788–818.

—— (2001). "The world's simplest grammars are creole grammars", *Linguistic Typology* 5: 125–166.

Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge: Cambridge University.

Mairal, Ricardo and Gil, Juana (2006a). "A first look at universals", in Ricardo Mairal and Juana Gil (eds.) (2006b), *Linguistic Universals*. Cambridge: Cambridge University, 1–45.

—— and —— (eds.) (2006b). *Linguistic Universals*. Cambridge: Cambridge University.

Maling, Joan (1984). "Non-clause-bounded reflexives in modern Icelandic", *Linguistics and Philosophy* 7: 211–241.

Mańczak, Witold (1958). "Tendences générales des changements analogiques", *Lingua* 7: 298–325 and 387–420.

—— (1980). "Laws of analogy", in Jacek Fisiak (ed.), *Historical Morphology*. The Hague: Mouton, 283–288.

—— (1987). *Frequenzbedingter unregelmäßiger Lautwandel in den germanischen Sprachen*. Wrocław: Ossolineum.

Mandilaras, Basil G. (1973). *The Verb in the Greek Non-Literary Papyri*. Athens: Hellenic Ministry of Culture and Sciences.

Marckwardt, Albert H. (1935). "Origin and extension of the voiceless preterit and the past participle inflections of the English irregular weak verb conjugation", in *Essays and Studies in English and Comparative Literature* (University of Michigan Publications, Language and Literature 13). Ann Arbor: University of Michigan, 151–328.

Martin, Samuel (1992). *A Reference Grammar of Korean*. Tokyo: Charles E. Tuttle.

Martins, Silvana and Martins, Valteir (1999). "Makú", in R. M. W. Dixon and Alexandra Y. Aikhenvald (eds.), *The Amazonian Languages*. Cambridge: Cambridge University, 251–267.

Masica, Colin P. (1991). *The Indo-Aryan Languages*. Cambridge: Cambridge University.

Matisoff, James A. (1973). *The Grammar of Lahu* (University of California Publications in Linguistics 75). Berkeley: University of California.

Mayerthaler, Willi (1981). *Morphologische Natürlichkeit*. Wiesbaden: Athenaion.

—— (1988). *Naturalness in Morphology.* Ann Arbor: Karoma. Eng. trans. of Mayerthaler (1981).

Meier-Brügger, Michael (1992). *Griechische Sprachwissenschaft*, 2 vols. Berlin: Walter de Gruyter.

Mielke, Jeff (2004). *The Emergence of Distinctive Features*. Columbus, Ohio:Ohio State University, Ph.D. thesis.

Mikheev, Andrei (1997). "Automatic rule induction for unknown-word guessing," *Computational Linguistics* 23: 405–423.

Milke, Wilhelm (1968). "Proto-Oceanic addenda", *Oceanic Linguistics* 7: 147–171.

Miller, Joanne (1994). "On the internal structure of phonetic categories: A progress report", *Cognition* 50: 271–285.

Mirčev, Kiril (1963). Istoričeska gramatika na bălgarskija ezik, 2nd edn. Sofia: Naukai Izkustvo.

—— (1978). *Istoričeska gramatika na bălgarskija ezik.* Sofia: Nauka i Izkustvo.

Molz, Hermann (1906). "Die Substantivflexion seit mittelhochdeutscher Zeit. II. Teil Neutra", *Paul und Braunes Beiträge zur Geschichte der deutschen Sprache und Literatur* 32: 277–392.

Morphy, Frances (1983). "Djapu, a Yolngu dialect", in R. M. W. Dixon and Barry J. Blake (eds.), *Handbook of Australian Languages, Volume III*. Amsterdam: Benjamins, 1–188.

Mossé, Fernand (1968). *A Handbook of Middle English*, trans. James A. Walker. Baltimore: Johns Hopkins.

Moulton, William G. (1947). "Juncture in Modern Standard German", *Language* 23: 212–216.

Mowrey, Richard and Pagliuca, William (1995). "The reductive character of articulatory evolution", *Rivista di Linguistica* 7: 37–124.

Mühlhäusler, Peter and Harré, Rom (1990). *Pronouns and People*. Oxford: Blackwell.

Nebieriʒe, Givi (1988). "Rogori sist'ema unda aγdges kartvelur puʒe-enaši—ergat'iuli tu nominat'iuri?" [What kind of system should be reconstructed in the Kartvelian proto-language—ergative or nominative?], *Macne* 2: 83–94.

Newman, John (1996). *Give: A Cognitive Linguistic Study*. Berlin: Mouton de Gruyter.

Newmeyer, Frederick J. (1998). *Language Form and Language Function*. Cambridge, Mass.: MIT.

—— (2004). "Against a parameter-setting approach to language variation", *Linguistic Variation Yearbook* 4: 181–234.

—— (2005). *Possible and Probable Languages: A Generative Perspective on Linguistic Typology.* Oxford: Oxford University.

Nichols, Johanna (1988). "On alienable and inalienable possession", in William Shipley (ed.), *In Honor of Mary Haas.* Berlin: Mouton de Gruyter, 475–521.

—— (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago.

Nichols, Johanna and Bickel, Balthasar (2005). "Possessive classification (alienable/inalienable possession)", in Martin Haspelmath, Matthew Dryer, Bernard Comrie, and David Gil (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University, 242–245.

Nichols, Johanna, Peterson, David A., and Barnes, Jonathan (2004). "Transitivizing and detransitivizing languages", *Linguistic Typology* 8: 149–211.

Norde, Muriel (2003). "[Review in Norwegian of] *Språk i endring. Indre norsk språkhistorie*", *Tijdschrift voor Skandinavistiek* 24: 271–282.

Ochs, Elinor, Schegloff, Emanuel A., and Thompson, Sandra A. (1996). "Introduction", in E. Ochs, S. Thompson, and E. Schegloff (eds.), *Interaction and Grammar* (Studies in Interactional Sociolinguistics 13). Cambridge: Cambridge University, 1–51.

Odden, David (2005). "The unnatural phonology of Zina Kotoko". Unpub. ms., The Ohio State University.

Ohala, John J. (1981). "The listener as a source of sound change", in Carrie S. Masek, Roberta A. Hendrick, and Mary Frances Miller (eds.), *Papers from the Parasession on Language and Behavior, Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, University of Chicago, 178–203.

—— (1990). "The phonetics and phonology of aspects of assimilation", in John Kingston and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University, 237–278.

—— (1993). "The phonetics of sound change", in Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*. London: Longman, 237–278.

Ohala, John J. and Kawasaki-Fukumori, Haruko (1997). "Alternatives to the sonority hierarchy for explaining the shape of morphemes", in Stig Eliasson and Ernst H. Jahr (eds.), *Studies for Einar Haugen*. Berlin: Mouton de Gruyter, 343–365.

Okell, John and Allott, Anna (2001). *Burmese/Myanmar Dictionary of Grammatical Forms*. London: Curzon.

Osborn, Henry A. (1962). *Warao Phonology and Morphology*. Indiana University, unpub. Ph.D. thesis.

Oudeyer, Pierre-Yves (2005). "The self-organization of speech sounds", *Journal of Theoretical Biology*, 233: 435–449.

Paradis, Carole and Prunet, Jean-François (eds.) (1991). *The Special Status of Coronals: Internal and External Evidence* (Phonetics and Phonology 2). San Diego: Academic.

Pätsch, Gertrud (1952). "Die georgische Aoristkonstruction", *Wissenschaftliche Zeitschrift der Humboldt-Universität Berlin (Gesellschafts- und sprachwissenschaftliche Reihe)* 1: 5–13.

Paul, Hermann (1880/1886). *Prinzipien der Sprachgeschichte*, 2nd edn. Halle: Niemeyer. Eng. trans. by Herbert Augustus Strong(1890/1970), *Principles of the History of Language*, New York: Macmillan.

—— (1880/1920). *Prinzipien der Sprachgeschichte*, 5th edn. Tübingen: Max Niemeyer.

Paul, Hermann, Wiehl, Peter, and Grosse, Siegfried (1989). *Mittelhochdeutsche Grammatik* (23rd edn.) . Tübingen: Max Niemeyer Verlag.

Pawley, Andrew (2001). "Proto Polynesian *-CIA", in Joel Bradshaw and Kenneth L. Rehg (eds.), *Issues in Austronesian Morphology: A Focusschrift for Byron W. Bender* (Pacific Linguistics 519). Canberra: Australian National University, 193–216.

Pawley, Andrew and Lane, Jonathan (1998). "From event sequence to grammar: Serial verb constructions in Kalam," in Anna Siewierska and Jae Jung Song (eds.), *Case, Typology and Grammar*. Amsterdam: Benjamins, 201–227.

Pawley, Andrew and Syder, Frances H. (2000). "The one clause at a time hypothesis", in Heidi Riggenbach (ed.), *Perspectives on Fluency*. Ann Arbor: University of Michigan, 163–199.

Payne, Doris L. (1986). "Basic constituent order in Yagua clauses: Implications for word order universals", in Desmond C. Derbyshire and Geoffrey K. Pullum (eds.), *Handbook of Amazonian Languages*, Volume 1. Berlin: Mouton de Gruyter, 440–468.

Payne, John R. (1979). "Transitivity and intransitivity in the Iranian languages of the U.S.S.R.", in Paul R. Clyne, William F. Hanks, and Carol L. Hofbauer (eds.), *The Elements: Papers from the Conference on Non-Slavic Languages of the USSR*. Chicago: Chicago Linguistic Society, 436–447.

Phillips, Betty S. (1984). "Word frequency and the actuation of sound change", *Language* 60: 320–342.

—— (2001). "Lexical diffusion, lexical frequency, and lexical analysis", in Joan Bybee and Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 123–136.

Pica, Pierre (1987). "On the nature of the reflexivization cycle", in Joyce McDonough and Bernadette Plunkett (eds.), *Proceedings of the Seventeenth Meeting of the North East Linguistic Society*. Amherst, Mass.: Graduate Linguistic Student Association, 483–499.

Pickett, Velma B. (1983). "Mexican Indian languages and Greenberg's 'Universals of Grammar' ", in Frederick B. Agard, Gerald Kelley, Adam Makkai, and Valerie Becker Makkai (eds.), *Essays in Honor of Charles F. Hockett*. Leiden: Brill, 530–551.

Pierrehumbert, Janet (2001). "Exemplar dynamics: Word frequency, lenition, and contrast", in Joan Bybee and Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins, 137–157.

—— (2002). "Word-specific phonetics", in Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology VII*. Berlin: Mouton de Gruyter, 101–139.

—— (2003). "Probabilistic phonology: Discrimination and robustness", in Rens Bod, Jennifer Hay, and Stefanie Jannedy (eds.), *Probabilistic Linguistics*. Cambridge, Mass.: MIT, 177–228.

Pike, Kenneth. (1947). "Grammatical prerequisites to phonemic analysis", *Word* 3: 155–172.

—— (1952). "More on grammatical prerequisites", *Word* 8: 106–121.

Pintzuk, Susan (2002). "Verb-object order in Old English", in David W. Lightfoot (ed.), *Syntactic Effects of Morphological Change*. Oxford: Oxford University, 276–299.

—— (2003). "Variationist approaches to syntactic change", in Brian D. Joseph and Richard D. Janda (eds.), *The Handbook of Historical Linguistics*. Oxford: Blackwell, 509–528.

Pintzuk, Susan, Tsoulas, George, and Warner, Anthony (eds.) (2000). *Diachronic Syntax: Models and Mechanisms*. Oxford: Oxford University.

Plank, Frans (ed.) (2003). *The Universals Archive*. Konstanz: Sprachwissenschaft, Universität Konstanz. http://ling.uni-konstanz.de/pages/proj/sprachbau.htm (accessed 20 Sept. 2003).

Postal, Paul M. (1966). "On so-called pronouns in English", in Francis P. Dinneen (ed.), *Report of the 17th Annual Round Table Meeting on Linguistics and Language Studies*.

Prince, Alan and Smolensky, Paul (1993). *Optimality Theory: Constraint Interaction in Generative Grammar* (Technical report RuCCS-TR-2). New Brunswick, NJ: Rutgers University Center for Cognitive Science. ROA-537, Rutgers Optimality Archive, http://roa.rutgers.edu/

Pullum, Geoffrey K. (1990). "Constraints on intransitive quasi-serial verb constructions in modern colloquial English", in Brian D. Joseph and Arnold M. Zwicky (eds.), *When Verbs Collide: Papers from the Ohio State Mini-Conference on Serial Verbs (Columbus, Ohio, May 26–27, 1990)* (Ohio State Working Papers in Linguistics 39). Columbus, Ohio:Ohio State University Department of Linguistics, 218–239.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, and Svartvik, Jan (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Raz, Shlomo (1983). *Tigre Grammar and Texts*. Malibu, Calif.: Undena Publications.

Reh, Mechthild (1985). *Die Krongo-Sprache (nìino mó-dì). Beschreibung, Texte, Wörterverzeichnis* (Kölner Beiträge zur Afrikanistik 12). Berlin: Reimer.

Reichard, Gladys A. (1925). "Wiyot Grammar and Texts", *University of California Publications in American Archaeology and Ethnology* 22: 1–213.

Reiter, Norbert (ed.) (1999). *Eurolinguistik: Ein Schritt in die Zukunft*. Wiesbaden: Harrassowitz.

Rice, Keren and Saxon, Leslie (2005). "Comparative Athapascan syntax", in Guglielmo Cinque and Richard S. Kayne (eds.), *Comparative Syntax*. Oxford: Oxford University, 698–774.

Rizzi, Luigi (1990). "On the anaphor-agreement effect", *Rivista di Linguistica* 2: 27–42.

Robins, R. H. (1958). *The Yurok Language: Grammar, Texts, Lexicon* (University of California Publications in Linguistics 15). Berkeley: University of California.

Rogava, G. (1975). "Nominat'iuri k'onst'rukciis mkone gardamavali zmnis genezisisatvis Kartvelur enebši" [On the genesis of transitive verbs that have the nominative construction in the Kartvelian languages], *C'elic'deuli* 2: 273–279.

Rogava, G. and Kerasheva, Z. I. (1966). *Grammatika adygejskogo jazyka*. Krasnodar: Krasnodarskoje knizhnoje izdatelstvo.

Romero-Figueroa, Andrés (1985). "OSV as the basic order in Warao", *Lingua* 66: 115–134.

Ross, Malcolm (1998). "Proto Oceanic phonology and morphology", in Malcolm Ross, Andrew Pawley, and Meredith Osmond (eds.), *The Lexicon of Proto Oceanic, Volume 1: Material Culture* (Pacific Linguistics C-152). Canberra: Australian National University, 14–35.

Rynell, Alarik (1948). *The Rivalry of Scandinavian and Native Synonyms in Middle English, especially* taken *and* nimen*, with an Excursion on* nema *and* taka *in Old Scandinavian* (Lund Studies in English XIII). Lund: C. W. K. Gleerup.

Sadler, Louisa (1997). "Clitics and the structure-function mapping", in Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG '97 Conference*. Stanford: CSLI. http://csli-publications.stanford.edu/LFG/2/lfg97.html

Sadock, Jerrold (1973). "Word-final devoicing in the development of Yiddish", in Braj B. Kachru, Robert Lees, Yakov Malkiel, Angelina Pietrangeli, and Sol Saporta (eds.), *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane*. Urbana: University of Illinois, 790–797.

Sakai, Hiromu (1998). "Feature checking and morphological merger", in David J. Silva (ed.), *Japanese/Korean Linguistics*, Volume 8. Stanford: CSLI, 189–202.

Sanderman, Alicia (2004). *A Corpus Study of Conditional Constructions in English*. Pittsburgh: Carnegie Mellon University College of Humanities and Social Sciences senior honors thesis.

Sanders, Gerald (1990). "On the analysis and implications of Maori verb alternations", *Lingua* 80: 150–198.

Šaniʒe, Ak'ak'i. (1953/1973). *Kartuli enis gramat'ik'is sapuʒvlebi* [Fundamentals of the Grammar of the Georgian Language]. Tbilisi: Universit'et'i.

Sansom, George (1928). *An Historical Grammar of Japanese*. Oxford: Clarendon.

Sapir, Edward (1915). "Notes on Judeo-German phonology", *Jewish Quarterly Review* 6: 231–266.

—— (1917). "Review of *Het passieve karakter van het verbum transitivum of van het verbum actionis in talen van Noord-Amerika*, by. C. C. Uhlenbeck", *International Journal of American Linguistics* 1: 82–86.

Saussure, Ferdinand de (1916/2005). *Course in General Linguistics*, trans. with annotations Roy Harris. London: Duckworth.

Schiller, Eric (1990). "The typology of serial verb constructions", in Michael Ziolkowski, Manuela Noske, and Karen Deaton (eds.), *Papers from the Twenty-Sixth Regional Meeting of the Chicago Linguistics Society*. Chicago: Chicago Linguistic Society, 393–406.

Schindler, Jochem (1974). "Fragen zum paradigmatischen Ausgleich", *Die Sprache* 20: 1–9.

Schmidt, Karl Horst (1966). "Tempora im Georgischen und in indogermanischen Sprachen", *Studia Caucasica* 2: 48–57.

—— (1973). "Transitive und intransitive", in Georges Redard (ed.), *Indogermanische und allgemeine Sprachwissenschaft*. Wiesbaden: Reichert, 9–24.

Schulze, Wolfgang (1988). *Studien zur Rekonstruktion des Lautstandes der südostkaukasischen (lezgischen) Grundsprache*. Bonn: Habilitationschrift, Universität Bonn.

—— (2004). "Review article, Alice C. Harris, *Endoclitics and the origins of Udi morphosyntax*", *Studies in Language* 28: 419–441.

—— (2005). "Towards a history of Udi", *International Journal of Diachronic Linguistics* 1.

Schwyzer, Eduard (1953). *Griechische Grammatik, Band 1: Allgemeiner Teil; Lautlehre; Wortbildung; Flexion*. München: C. H. Beck.

Sebba, Mark (1994). "Serial verbs", in R. E. Asher (ed.), *The Encyclopedia of Languages and Linguistics*. Oxford: Pergamon, 3858–3861.

Seiter, William (1980). *Studies in Niuean Syntax*. New York: Garland.

Sells, Peter (1995). "Korean and Japanese morphology from a lexical perspective", *Linguistic Inquiry*: 277–325.

Šerozia, Revaz (1980). "P'ot'encialisis k'at'egoria da mastan dak'avširebuli zogi sak'itxi kartvelur enebši" [The category of potential and some questions related to it in the Kartvelian languages], in G. Bedošvili and B. Jorbenaʒe (eds.), *Nark'vevebi iberiul-k'avk'asiur enata morpologiidan*. Tbilisi: Mecniereba, 119–126.

Shih, Chilin (2005). "Understanding phonology by phonetic implementation", *INTERSPEECH-2005*, 2469–2472.

Siemund, P. (2001). "Interrogative constructions", in Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), *Language Typology and Language Universals: An International Handbook*, Volume 2. Berlin: Walter de Gruyter, 1010–1028.

Siewierska, Anna (2005). "Third person zero of verbal person marking", in Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie (eds.), *The World Atlas of Language Structures*. Oxford: Oxford University, 418–421.

Silverstein, Michael (1976). "Hierarchy of features and ergativity", in R. M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*. Canberra: Australian Institute of Aboriginal Studies, 112–171.

Smith, Neilson (1973). *The Acquisition of Phonology*. Cambridge: Cambridge University.

Sohn, Ho-Min (1999). *The Korean Language*. Cambridge: Cambridge University.

Sohn, Hyang-Sook (n.d.). "On the role of paradigm uniformity in the lexicon'. Unpub. ms., Kyungpook National University.

Sommer, Bruce (1969). *Kunjen Phonology: Synchronic and Diachronic* (Pacific Linguistics B-11). Canberra: Australian National University.

—— (1970). "An Australian language without CV syllables", *International Journal of American Linguistics* 36: 57–58.

Sommerstein, Alan (1982). [Aristophanes'] *Clouds.* Warminster: Aris and Phillips.

Sportiche, Dominique (1996). "Clitic constructions", in Johan Rooryck and Laurie Zaring (eds.), *Phrase Structure and the Lexicon.* Bloomington: Indiana University Linguistics Club, 213–276.

Sproat, Richard (1985). "Welsh syntax and VSO structure", *Natural Language and Linguistic Theory* 3: 173–216.

Stampe, David (1973). *A Dissertation on Natural Phonology*. University of Chicago, Ph.D. thesis.

—— (1979). *A Dissertation on Natural Phonology*. New York: Garland. Published version of Stampe (1973).

Stassen, Leo (2000). "AND-languages and WITH-languages", *Linguistic Typology* 4: 1–54.

Steels, Luc (1997). "Self-organizing vocabularies", in Christopher G. Langton and Katsunori Shimohara (eds.), *Proceedings of the Fifth International Workshop on Artificial Life: Synthesis and Simulation of Living Systems*. Cambridge, Mass.: MIT, 179–84.

—— (2000). "Language as a complex adaptive system", in Marc Schoenauer (ed.), *Parallel Problem Solving from Nature—PPSN VI: Sixth International Conference, Paris, France, September 2000, Proceedings*. Berlin: Springer, 17–26.

Steriade, Donca (1999a). "Alternatives to syllable-based accounts of consonantal phonotactics", in Osamu Fujimura, Brian D. Joseph, and Bohumil Palek (eds.), *Proceedings of the 1998 Linguistics and Phonetics Conference*. Prague: Karolinum, 205–242.

—— (1999b). "Lexical conservatism in French adjectival liaison", in Jean-Marc Authier, Barbara E. Bullock, and Lisa A. Reed (eds.), *Formal Perspectives on Romance Linguistics*. Amsterdam: Benjamins, 243–270.

—— (2000). "Paradigm uniformity and the phonetics-phonology boundary", in Michael Broe and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon.* Cambridge: Cambridge University, 313–334.

—— (in press). "The phonology of perceptibility effects: The P-map and its consequences for constraint organization", in Kristin Hanson and Sharon Inkelas (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge, Mass.: MIT.

Stilo, Donald (1987). "Ambipositions as an areal response: The case study of the Iranian zone", in Elena Bashir, Madhav M. Deshpande, and Peter E. Hook (eds.), *Select Papers from SALA-7: South Asian Languages Analysis Roundtable Conference, Held in Ann Arbor, Michigan, May 17–19, 1985*. Bloomington: Indiana University Linguistics Club, 308–334.

—— (2005). "Iranian as buffer zone between the universal typologies of Turkic and Semitic", in Éva Ágnes Csató, Bo Isaksson, and Carina Jahani (eds.), *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*. London: RoutledgeCourzon, 35–63.

—— (forthcoming). *The Araxes Sprachbund.*

Stresemann, Erwin (1927). "Die Lauterscheinungen in den Ambonischen Sprachen", *Zeitschrift für Eingeborenen-Sprachen*, Supplement 10. Berlin: Dietrich Reimer.

Stroomer, Harry (1987). *A Comparative Study of Three Southern Oromo Dialects in Kenya*. Hamburg: Helmut Buske.

Studdert-Kennedy, Michael (1987). "The phoneme as a perceptuomotor structure", in Alan Allport, Donald G. MacKay, Wolfgang Prinz, and Eckart Scheerer (eds.), *Language, Perception, and Production: Relationships between Listening, Speaking, Reading, and Writing*. New York: Academic, 67–84.

Studdert-Kennedy, Michael (1988). "The particulate origins of language generativity: From syllable to gesture", in James Hurford, Michael Studdert-Kennedy, and Chris Knight (eds.), *Approaches to the Evolution of Language*. Cambridge: Cambridge University, 202–221.

Sun, Chaofen (1996). *Word-Order Change and Grammaticalization in the History of Chinese.* Stanford: Stanford University.

Sutton, Peter (1978).*Wik: Aboriginal Society, Territory and Language at Cape Keerweer*. University of Queensland, Ph.D. thesis.

Sweetser, Eve E. (1990). *From Etymology to Pragmatics*. Cambridge: Cambridge University.

Teeter, Karl V. (1964). *The Wiyot Language* (University of California Publications in Linguistics 37). Berkeley: University of California.

Tesar, Bruce and Prince, Alan (2007). "'Using phototactics to learn phonological alternations"', in Jonathan E. Cihlar, Amy L. Franklin, David W. Kaiser, and Irene Kimbara (eds.), *CLS 39: 2. The Panels: Papers from the 39th Annual Meeting of The Chicago Linguistic Society*, 209–237.

Thompson, Chad (1996). "On the grammar of body parts in Koyukon Athabaskan", in Hilary Chappell and William McGregor (eds.), *The Grammar of Inalienability.* Berlin: Mouton de Gruyter, 651–676.

Tiersma, Peter Meijes (1982). "Local and general markedness", *Language* 58: 832–849.

Timberlake, Alan (1975). "Hierarchies in the genitive of negation", *Slavic and East European Journal* 19: 123–138.

—— (2004). *A Reference Grammar of Russian*. Cambridge: Cambridge University.

Toivonen, Ida (2001). *Non-Projecting Words: Evidence from Verbal Particles in Swedish*. Stanford University, Ph.D. thesis.

Trask, R. L. (1996). *Historical Linguistics.* London: Arnold.

Travis, Lisa (1984). *Parameters and Effects of Word Order Variation*. MIT, Ph.D. thesis.

Trudgill, Peter (1983). *On Dialect: Social and Geographical Perspectives*. Oxford: Blackwell.

—— (1989). "Contact and evolution in linguistic change", in Leiv Egil Breivik and Ernst Håkon Jahr (eds.), *Language Change: Contributions to the Study of its Causes*. Berlin: Mouton de Gruyter, 227–237.

—— (1996). "Dialect typology: Isolation, social network and phonological structure", in Gregory R. Guy, Crawford Feagin, Deborah Schiffrin, and John Baugh (eds.), *Towards a Social Science of Language: Papers in Honor of William Labov, Volume 1: Variation and Change in Language and Society*. Amsterdam: Benjamins, 3–21.

Ultan, Russell (1978). "Interrogative systems" in Joseph Greenberg, Charles Ferguson, and Edith Moravcsik (eds.) *Universals of Human Language, Volume 4: Syntax*. Stanford: Stanford University Press 211–248.

Vaquero, Antonio (1965). *Idioma Warao: Morphologia, sintaxis, literatura*. Caracas: Editorial Sucre.

Vaux, Bert (2002). "Consonant epenthesis and the problem of unnatural phonology'. Handout of talk presented at the Yale University Linguistics Colloquium.

Vaux, Bert and Samuels, Bridget (2005). "Laryngeal markedness and aspiration", *Phonology* 22: 395–436.

Vennemann, Theo (1972a). "Phonetic analogy and conceptual analogy", in Theo Vennemann and Terence H. Wilbur (eds.), *Schuchhardt, the Neogrammarians, and the Transformational*

Theory of Phonological Change: Four Essays by Hugo Schuchhardt, Theo Vennemann, Terence H. Wilbur (Linguistische Forschungen 26). Frankfurt am Main: Athenäum, 115–179.

Vennemann, Theo (1972b). "Rule inversion", *Lingua* 29: 209–242.

Verma, Manindra and Mohanan, K. P. (1990). "Introduction to the experiencer subject construction", *Experiencer Subjects in South Asian Languages*. Stanford: CSLI, 1–12.

Verner, Karl (1875). "Eine Ausnahme der ersten Lautverschiebung", [Kuhn's] *Zeitschrift für vergleichende Sprachforschung* 23: 97–130.

Vikner, Sten (1995). V*erb Movement and Expletive Movement in the Germanic Languages*. Oxford: Oxford University.

Watkins, Calvert (1963). "Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb", *Celtica* 6: 1–49.

—— (1964). "Preliminaries to the reconstruction of Indo-European sentence structure", in Horace Gray Lunt (ed.), *Proceedings of the Ninth International Congress of Linguists, Cambridge, Mass., August 27–31, 1962*. The Hague: Mouton, 1035–1045.

Wedel, Andrew (2004). *Self-organization and Categorical Behavior in Phonology*. University of California, Santa Cruz, Ph.D. thesis.

—— (2006). "Exemplar models, evolution and language change", *Linguistic Review* 23: 247–274.

Weir, E. M. Helen (1984). *A negação e outros tópicos da grámatica Nadëb*. Universidade Estadual de Campinas, M.A. thesis.

Wexler, Kenneth (1998). "Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage", *Lingua* 106: 23–79.

Wheeler, Benjamin Ide (1887). *Analogy and the Scope of its Application in Language.* New York: J. Wilson.

Whitman, John and Paul, Waltraud (2005). "Reanalysis and conservancy of structure in Chinese", in Montserrat Batllori, Maria-Lluïsa Hernanz, Carme Picallo and Francesc Roca (eds.), *Grammaticalization and Parametric Variation*. Oxford: Oxford University, 82–94.

Wierzbicka, Anna (1981). "Case marking and human nature", *Australian Journal of Linguistics* 1: 43–81.

Windfuhr, Gernot (1987). "Persian", in Bernard Comrie (ed.), *The World's Major Languages*. New York: Oxford University, 523–546.

Wissing, Daan and Zonneveld, Wim (1996). "Final devoicing as a robust phenomenon in second language acquisition: Tswana, English and Afrikaans", *South African Journal of Linguistics*, Supplement 34, 3–24.

Woolford, Ellen (1999). "More on the anaphor agreement effect", *Linguistic Inquiry* 30: 257–287.

—— (to appear). "Differential subject marking at argument structure, syntax, and PF", in Helen de Hoop and Peter de Swart (eds.), *Differential Subject Marking* (Studies in Natural Language and Linguistic Theory). Dordrecht: Kluwer.

Wray, Alison (2004). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University.

Wright, Saundra Kimberly (2001). *Internally Caused and Externally Caused Change of State Verbs.* Northwestern University, Ph.D. thesis.

Wunderlich, Dieter (1997). "Cause and the structure of verbs", *Linguistic Inquiry* 28: 27–68.

Xajdakov, S. M. (1986). "Logičeskoe udarenie i ílenenie predloženija (dagestanskie dannye)", *Aktual'nye problemy dagestansko-nakhskogo jazykoznanija*. Maxačkala: In-t istorii, 79–96.

Xrakovskij, Viktor S. (ed.) (2001). *Typology of Imperative Constructions.* Munich: LINCOM.

Yavaş, Mehmet S. (1994). "Final stop devoicing in interlanguage", in Mehmet S. Yavaş (ed.), *First and Second Language Phonology*. San Diego: Singular, 267–282.

Yoon, Kyuchul, Beckman, Mary, and Brew, Chris (2002). "Letter-to-sound rules for Korean", *Proceedings of 2002 IEEE Workshop on Speech Synthesis, 11–13 September 2002, Santa Monica, CA*. Piscataway, NJ: IEEE, 47–50.

Yu, Alan C. L. (2003). *The Morphology and Phonology of Infixation*. University of California, Berkeley, Ph.D. Thesis.

—— (2004). "Explaining final obstruent voicing in Lezgian: Phonetics and history", *Language* 80: 73–97.

Yusuf, Ore (1986). *Verb Phrase Serialization in Yoruba in Discourse Perspective.* University of California, Los Angeles, Ph.D. thesis.

Zipf, George K. (1935). *The Psycho-Biology of Language: An Introduction to Dynamic Philology.* Boston: Houghton Mifflin. Repub. (1965), Cambridge, Mass.: MIT.

—— (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Cambridge, Mass.: Addison-Wesley.

Zoe Wu, Xiu-Zhi (2004). *Grammaticalization and Language Change in Chinese.* London: RoutledgeCurzon.

Zorell, Franz (1930). *Grammatik zur altgeorgischen Bibelübersetzung*. Rome: Pontificium Institutum Biblicum.

Zúñiga, Fernando (2006). *Deixis and Alignment: Inverse Systems in Indigenous Languages of the Americas* (Typological Studies in Language 70). Amsterdam: Benjamins.

Zuraw, Kie (2000). *Patterned Exceptions in Phonology*. University of California, Los Angeles, Ph.D. thesis.

Zwicky, Arnold M. and Pullum, Geoffrey K. (1983). "Cliticization vs. inflection: English *n't*", *Language* 59: 502–513.

This page intentionally left blank

INDEX